

PERSPECTIVES IN LOGIC

Petr Hájek
Pavel Pudlák

METAMATHEMATICS
OF FIRST-ORDER
ARITHMETIC



ASL

CAMBRIDGE

Metamathematics of First-Order Arithmetic

Since their inception, the Perspectives in Logic and Lecture Notes in Logic series have published seminal works by leading logicians. Many of the original books in the series have been unavailable for years, but they are now in print once again.

This volume, the 3rd publication in the Perspectives in Logic series, is a much-needed monograph on the metamathematics of first-order arithmetic. The authors pay particular attention to subsystems (fragments) of Peano arithmetic and give the reader a deeper understanding of the role of the axiom schema of induction and of the phenomenon of incompleteness.

The reader is only assumed to know the basics of mathematical logic; these are reviewed in the preliminaries. Part A develops parts of mathematics and logic in various fragments. Part B is devoted to incompleteness. Finally, Part C studies systems that have the induction schema restricted to bounded formulas (bounded arithmetic).

PETR HÁJEK works in the Institute of Computer Science at the Academy of Sciences of the Czech Republic, Prague.

PAVEL PUDLÁK works in the Mathematical Institute at the Academy of Sciences of the Czech Republic, Prague.

PERSPECTIVES IN LOGIC

The *Perspectives in Logic* series publishes substantial, high-quality books whose central theme lies in any area or aspect of logic. Books that present new material not now available in book form are particularly welcome. The series ranges from introductory texts suitable for beginning graduate courses to specialized monographs at the frontiers of research. Each book offers an illuminating perspective for its intended audience.

The series has its origins in the old *Perspectives in Mathematical Logic* series edited by the Ω -Group for “Mathematische Logik” of the Heidelberger Akademie der Wissenschaften, whose beginnings date back to the 1960s. The Association for Symbolic Logic has assumed editorial responsibility for the series and changed its name to reflect its interest in books that span the full range of disciplines in which logic plays an important role.

Arnold Beckmann, Managing Editor

Department of Computer Science, Swansea University

Editorial Board:

Michael Benedikt

Department of Computing Science, University of Oxford

Elisabeth Bouscaren

CNRS, Département de Mathématiques, Université Paris-Sud

Steven A. Cook

Computer Science Department, University of Toronto

Michael Glanzberg

Department of Philosophy, University of California Davis

Antonio Montalban

Department of Mathematics, University of Chicago

Simon Thomas

Department of Mathematics, Rutgers University

For more information, see www.aslonline.org/books_perspectives.html

PERSPECTIVES IN LOGIC

Metamathematics of First-Order Arithmetic

PETR HÁJEK

Academy of Sciences of the Czech Republic

PAVEL PUDLÁK

Academy of Sciences of the Czech Republic



ASSOCIATION FOR SYMBOLIC LOGIC



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107168411

10.1017/9781316717271

First edition © 1998 Springer-Verlag Berlin Heidelberg

This edition © 2016 Association for Symbolic Logic under license to
Cambridge University Press.

Association for Symbolic Logic

Richard A. Shore, Publisher

Department of Mathematics, Cornell University, Ithaca, NY 14853

<http://www.aslonline.org>

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

A catalogue record for this publication is available from the British Library.

ISBN 978-1-107-16841-1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy
of URLs for external or third-party Internet Web sites referred to in this publication
and does not guarantee that any content on such Web sites is, or will remain,
accurate or appropriate.

Dedicated to our wives, Marie and Věra

Preface to the Series

Perspectives in Mathematical Logic

(Edited by the "Ω-group for Mathematical Logic" of the
Heidelberger Akademie der Wissenschaften)

On Perspectives. *Mathematical logic* arouse from a concern with the nature and the limits of rational or mathematical thought, and from a desire to systematise the modes of its expression. The pioneering investigations were diverse and largely autonomous. As time passed, and more particularly in the last two decades, interconnections between different lines of research and links with other branches of mathematics proliferated. The subject is now both rich and varied. It is the aim of the series to provide, as it were, maps of guides to this complex terrain. We shall not aim at encyclopaedic coverage: nor do we wish to prescribe, like Euclid, a definitive version of the elements of the subject. We are not committed to any particular philosophical programme. Nevertheless we have tried by critical discussion to ensure that each book represents a coherent line of thought; and that, by developing certain themes, it will be of greater interest than a mere assemblage of results and techniques.

The books in the series differ in level: some are introductory, some highly specialised. They also differ in scope: some offer a wide view of an area, others present a single line of thought. Each book is, at its own level, reasonably self-contained. Although no book depends on another as prerequisite, we have encouraged authors to fit their books with other planned volumes, sometimes deliberately seeking coverage of the same material from different points of view. We have tried to attain a reasonable degree of uniformity of notation and arrangement. However, the books in the series are written by individual authors, not by the group. Plans for books are discussed and argued about at length. Later, encouragement is given and revisions suggested. But it is the authors who do the work; if, as we hope, the series proves of value, the credit will be theirs.

History of the Ω-Group. During 1968 the idea of an integrated series of monographs on mathematical logic was first mooted. Various discussions led to a meeting at Oberwolfach in the spring of 1969. Here the founding members of the group (R.O. Gandy, A. Levy, G.H. Müller, G. Sacks, D.S. Scott) discussed the project in earnest and decided to go ahead with it. Professor F.K. Schmidt and Professor Hans Hermes gave us encouragement and support. Later Hans

Hermes joined the group. To begin with all was fluid. How ambitious should we be? Should we write the books ourselves? How long would it take? Plans for authorless books were promoted, savaged and scrapped. Gradually there emerged a form and a method. At the end of an infinite discussion we found our name, and that of the series. We established our centre in Heidelberg. We agreed to meet twice a year together with authors, consultants and assistants, generally in Oberwolfach. We soon found the value of collaboration: on the one hand the permanence of the founding group gave coherence to the over all plans; on the other hand the stimulus of new contributors kept the project alive and flexible. Above all, we found how intensive discussion could modify the authors' ideas and our own. Often the battle ended with a detailed plan for a better book which the author was keen to write and which would indeed contribute a perspective.

Oberwolfach, September 1975

Acknowledgements. In starting our enterprise we essentially were relying on the personal confidence and understanding of Professor Martin Barner of the Mathematisches Forschungsinstitut Oberwolfach, Dr. Klaus Peters of Springer-Verlag and Dipl.-Ing. Penschuck of the Stiftung Volkswagenwerk. Through the Stiftung Volkswagenwerk we received a generous grant (1970-1973) as an initial help which made our existence as a working group possible.

Since 1974 the Heidelberger Akademie der Wissenschaften (Mathematisch-Naturwissenschaftliche Klasse) has incorporated our enterprise into its general scientific program. The initiative for this step was taken by the late Professor F.K. Schmidt, and the former President of the Academy, Professor W. Doerr.

Through all the years, the Academy has supported our research project, especially our meetings and the continuous work on the Logic Bibliography, in an outstandingly generous way. We could always rely on their readiness to provide help wherever it was needed.

Assistance in many various respects was provided by Drs. U. Felgner and K. Gloede (till 1975) and Drs. D. Schmidt and H. Zeitler (till 1979). Last but not least, our indefatigable secretary Elfriede Ihrig was and is essential in running our enterprise.

We thank all those concerned.

Heidelberg, September 1982

R.O. Gandy

H. Hermes

A. Levy

G.H. Müller

G. Sacks

D.S. Scott

Authors' Preface

After having finished this book on the metamathematics of first order arithmetic, we consider the following aspects of it important: first, we pay much attention to subsystems (fragments) of the usual axiomatic system of first order arithmetic (called Peano arithmetic), including weak subsystems, i.e. so-called bounded arithmetic and related theories. Second, before discussing proper metamathematical questions (such as incompleteness) we pay considerable attention to positive results, i.e. we try to develop naturally important parts of mathematics (notably, some parts of set theory, logic and combinatorics) in suitable fragments. Third, we investigate two notions of relative strength of theories: interpretability and partial conservativity. Fourth, we offer a systematic presentation of relations of bounded arithmetic to problems of computational complexity.

The need for a monograph on metamathematics of first order arithmetic has been felt for a long time; at present, besides our book, at least two books on this topic are to be published, one written by R. Kaye and one written by C. Smoryński. We have been in contacts with both authors and are happy that the overlaps are reasonably small so that the books will complement each other.

This book consists of a section of preliminaries and of three parts: A – Positive results on fragments, B – Incompleteness, C – Bounded arithmetic. Preliminaries and parts A, B were written by P. H., part C by P. P. We have tried to keep all parts completely compatible.

The reader is assumed to be familiar with fundamentals of mathematical logic, including the completeness theorem and Herbrand's theorem; we survey the things assumed to be known in the Preliminaries, in order to fix notation and terminology.

Acknowledgements. Our first thanks go to the members of the Ω -group for the possibility of publishing the book in the series Perspectives in mathematical logic and especially to Professor Gert H. Müller, who invited P. H. to write a monograph with the present title, agreed with his wish to write the book jointly with P. P. and continuously offered every possible help. We

are happy to recognize that we have been deeply influenced by Professor Jeff Paris. Soon after the famous independence results of Paris, Kirby and Harrington, Jeff Paris repeatedly visited Prague and gave talks about the research of his Manchester group. Since then, he has come to Prague many times and we always learn much from him. On various occasions we met other mathematicians working in this field (Adamowicz, Buss, Clote, Dimitracopoulos, Feferman, Kaye, Kossak, Kotlarski, Lindström, Montagna, Ressayre, Simpson, Smoryński, Solovay, Takeuti, Wilkie, Woods and others) and many of them visited Czechoslovakia. Discussions with them and preprints of their papers have been an invaluable source of information for us. We have profited extremely much from our colleagues J. Krajíček and V. Švejdar and other members of our Prague seminar. The Mathematical Institute of the Czechoslovak Academy of Sciences has been a good working place. Several people have read parts of the manuscript and suggested important improvements. Our thanks especially to Peter Clote, William Eldridge, Richard Kaye, Juraj Hromkovič and Jiří Sgall for their help. Mrs. K. Trojanová and Mrs. D. Berková helped us considerably with typing; and D. Harmanec provided valuable technical help with the preparation of the bibliography on a computer. Last but not least, our families have got used to sacrifice for our scientific work. They deserve our most cordial thanks.

November 1990

Petr Hájek
Pavel Pudlák

Table of Contents

Introduction	1
Preliminaries	5
(a) Some Logic	5
(b) The Language of Arithmetic, the Standard Model	12
(c) Beginning Arithmetization of Metamathematics	20
PART A	
CHAPTER I	
Arithmetic as Number Theory, Set Theory and Logic	27
Introduction	27
1. Basic Developments; Partial Truth Definitions	28
(a) Properties of Addition and Multiplication, Divisibility and Primes	28
(b) Coding Finite Sets and Sequences; the Theory $I\Sigma_0(exp)$	37
(c) Provably Recursive Functions; the Theory $I\Sigma_1$	44
(d) Arithmetization of Metamathematics: Partial Truth Definitions	50
2. Fragments of First-Order Arithmetic	61
(a) Induction and Collection	61
(b) Further Principles and Facts About Fragments	67
(c) Finite Axiomatizability; Partial Truth Definitions for Relativized Arithmetical Formulas	77
(d) Relativized Hierarchy in Fragments	81
(e) Axiomatic Systems of Arithmetic with No Function Symbols	86
3. Fragments and Recursion Theory	89
(a) Limit Theorem	89
(b) Low Basis Theorem	91
(c) Infinite Δ_1 Subsets	95
(d) Matiyasevič's Theorem in $I\Sigma_1$	97
4. Elements of Logic in Fragments	98
(a) Arithmetizing Provability	98

(b) Arithmetizing Model Theory	102
(c) Applications to Arithmetic	105
CHAPTER II	
Fragments and Combinatorics	111
1. Ramsey's Theorems and Fragments	111
(a) Statement of Results	111
(b) Proofs (of 1.5, 1.7, 1.9)	115
(c) Proofs (of 1.6, 1.8, 1.10)	118
2. Instances of the Paris-Harrington Principle and Consistency Statements	121
(a) Introduction and Statement of Results	121
(b) Some Combinatorics	122
(c) Proof of $Con^*(I\Sigma_u^* + Tr(\Pi_1^*)) \rightarrow (PH)_u$ (for $u \geq 1$)	124
(d) Strong Indiscernibles	125
(e) Final Considerations	129
3. Schwichtenberg-Wainer Hierarchy and α -large Sets	132
(a) Ordinals in $I\Sigma_1$	133
(b) Transfinite Induction and Fragments	138
(c) α -large Sets in $I\Sigma_1$	139
(d) Schwichtenberg-Wainer Hierarchy	140
PART B	
CHAPTER III	
Self-Reference	147
1. Preliminaries	148
(a) Interpretability and Partial Conservativity	148
(b) Theories Containing Arithmetic; Sequential Theories; PA and ACA_0	150
(c) Numerations and Binumerations	155
2. Self-Reference and Gödel's Theorems, Reflexive Theories	158
(a) Existence of Fixed Points	158
(b) Gödel's First Incompleteness Theorem and Related Topics	160
(c) Gödel's Second Incompleteness Theorem	163
(d) Pure Extensions of PA	168
(e) Interpretability in Pure Extensions of PA	169
3. Definable Cuts	171
(a) Definable Cuts and Their Properties	172
(b) A Strong Form of Gödel's Second Incompleteness Theorem	173
(c) Herbrand Provability and Herbrand Consistency	179
(d) Cuts and Interpretations	186
4. Partial Conservativity and Interpretability	189
(a) Some Prominent Examples	190

(b) General Theorems on Partial Conservativity; Some Fixed-Point Theorems	195
(c) Applications, Mainly to Interpretability	206

CHAPTER IV

Models of Fragments of Arithmetic	213
1. Some Basic Constructions	214
(a) Preliminaries	214
(b) Definable Ultrapower of the Standard Model	216
(c) On Submodels and Cuts	218
(d) Models for the Hierarchy	220
(e) Elementary End Extensions	227
(f) A Conservation Result	230
2. Cuts in Models of Arithmetic with a Top	232
(a) Arithmetic with a Top and Its Models	232
(b) Cuts	234
(c) Extendable, Restrainable and Ramsey Cuts	236
(d) Satisfaction in Finite Structures with an Application to Models of $I\Sigma_1$	241
3. Provably Recursive Functions and the Method of Indicators	245
(a) Provably Recursive Functions, Envelopes	245
(b) Indicators and Paris Sequences	247
(c) Paris Sequences of the First Kind	250
(d) Paris Sequences of the Second Kind	253
(e) Further Consequences	257
4. Formalizing Model Theory	258
(a) Some Results on Satisfaction and Consistency	259
(b) A Conservation Result in $I\Sigma_1$	260
(c) Appendix: Another Conservation Result	263

PART C

CHAPTER V

Bounded Arithmetic	267
1. A Survey of Weak Fragments of Arithmetic	268
(a) Fragments of Arithmetic	268
2. A Brief Introduction to Complexity Theory	276
(a) Time and Space Complexity Classes	277
(b) Nondeterministic Computations	279
(c) Degrees and NP -completeness	280
(d) Oracle Computations	282
(e) The Linear Time Hierarchy and the Polynomial Hierarchy	283
(f) Nepomnjaščij's Theorem	285
(g) The Diagonal Method for Separating Complexity Classes	288

3. Exponentiation, Coding Sequences and Formalization of Syntax in $I\Sigma_0$	294
(a) Introduction	294
(b) Sets and Sequences	295
(c) The Exponentiation Relation	299
(d) Developing $I\Sigma_0 + \Omega_1$	303
(e) The Number of Ones in a Binary Expansion	304
(f) Coding Sequences	309
(g) Syntactical Concepts	312
(h) Formalizations Based on Context-Free Grammars	315
4. Witnessing Functions	320
(a) Introduction	320
(b) Fragments of Bounded Arithmetic	320
(c) Definability of Turing Machine Computations in Fragments of Bounded Arithmetic	330
(d) Witnessing Functions	337
(e) On the Finite Axiomatizability of Bounded Arithmetic	350
5. Interpretability and Consistency	360
(a) Introduction	360
(b) Truth Definitions for Bounded Formulae	361
(c) An Interpretation of $I\Sigma_0$ in Q	366
(d) Cut-Elimination and Herbrand's Theorem in Bounded Arithmetic	371
(e) The Π_1 Theorems of $I\Sigma_0 + Exp$	380
(f) Incompleteness Theorems	386
(g) On the Limited Use of Exponentiation	393
Bibliographical Remarks and Further Reading	397
Bibliography	409
Index of Terms	455
Index of Symbols	459

Introduction

People have been interested in natural numbers since forever. The ancient mathematicians knew and used the principle of *descente infinie*, which is a form of mathematical induction. The principle is as follows: if you want to show that no number has the property φ , it suffices to show that for each number n having the property φ there is a smaller number $m < n$ having the property φ . (If there were a number having φ we could endlessly find smaller and smaller numbers having φ , which is absurd.) The Greeks used the principle for a proof of incommensurability of segments. The principle was rediscovered in modern times by P. Fermat (1601–1665). The principle of mathematical induction itself (if 0 has the property φ and for each number n having φ also $n+1$ has φ then all numbers have φ) seems to have been first used by B. Pascal (1623–1662) in a proof concerning his triangle. A general formulation appears in a work of J. Bernoulli (1654–1705). (Our source is [Meschkowski 78–81].)

In 1861 Grassman published his *Lehrbuch der Arithmetik*; in our terms, he defines integers as an ordered integrity domain in which each non-empty set of positive elements has a least element. In 1884 Frege's book *Grundlagen der Arithmetik* was published. We can say that Frege's natural numbers are classes; each such class consists of all sets of a certain fixed finite cardinality. (Frege speaks of concepts, not of classes.) The famous Dedekind's work *Was sind und was sollen die Zahlen* appears in 1888. Dedekind's natural numbers are defined as a set N together with an element $1 \in N$ a one-one mapping f of N into itself such that 1 is not in the range of f and N is the smallest set containing 1 and closed under f . Dedekind and Frege agreed that arithmetic is a part of logic, but differed in their opinions on what logic is. They both used the same main device: a one-one mapping and closedness under that mapping.

Dedekind was not interested in finding a formal deductive system for natural numbers; this was the main aim of Peano's investigation of natural numbers (*Arithmetices principia nova methoda exposita*, 1889). Peano's axiom system (taken over from Dedekind, who had it from Grassman) is, in our terminology, second order: it deals with numbers and sets of numbers. Nowadays

it is usual to call the first-order axiomatic arithmetic Peano arithmetic; this terminology was probably introduced by Tarski (personal communication by G.H. Müller). Whitehead and Russell published their *Principia mathematica* in 1908; the book also includes a formalization of arithmetic.

Hilbert formulated his programme as follows: *unsere üblichen Methoden der Mathematik samt und sonders als widerspruchsfrei zu erkennen* (to show that our usual methods of mathematics are free from contradictions in their whole). [Hilbert-Bernays 34, Zur Einleitung]. This should have been shown by finitary methods forming a proper part of arithmetic [ibid., p. 42]. Gödel's famous incompleteness results [Gödel 31] showed Hilbert's program in its original formulation to be unrealistic (even if Hilbert denies this in his Einleitung); but it has remained an important source of inspiration for proof theory, see [Kreisel 68]. Related work from the thirties by Tarski (undefinability of truth in arithmetic), Church (undecidability of arithmetic) and Rosser (elimination of the assumption of ω -consistency) is well known. In modern texts these results are proved using the well-known diagonalization (or self-reference) lemma, which is already implicit in Gödel's proof. This lemma first appeared explicitly in [Carnap 34], but, surprisingly, it was neglected by many authors for a long time. Feferman's paper [Feferman 60] is a fundamental paper for modern study of arithmetization of metamathematics. But it is also necessary to mention Volume II of Hilbert-Bernays's monograph [Hilbert-Bernays 39], containing a detailed exposition of arithmetization including the arithmetized completeness theorem. Early results following Feferman's Arithmetization were obtained by Montague, Shepherdson and others. In the sixties, Feferman and Montague worked on a monograph devoted to the arithmetization of metamathematics, but unfortunately the book has never been finished. [Smoryński 81-fifty] is a very readable survey of the development of self-reference.

Non-standard models of arithmetic were first constructed by Skolem [Skolem 34]; in present terms, he used the method of definable ultrapower. In 1952 Ryll-Nardzewski proved that Peano arithmetic PA (first order!) is not finitely axiomatizable. Specker and McDowell showed in 1959 that each (countable) model of PA has an elementary end-extension. Rabin [61] showed that PA is not axiomatizable by any axiom system of bounded quantifier complexity. Further important results were obtained by Friedman, Gaifman and Paris in the early seventies. [Smoryński 82] is a very readable treatise of development of model theory of arithmetic (up to the early eighties).

A result of fundamental importance was obtained by Paris in 1977: he found an arithmetical statement with a clear combinatorial meaning which is true but unprovable in PA; moreover, he was able to show the unprovability by model-theoretical means, without any use of self-reference. His proof used a new method, called the method of indicators, developed by Paris and Kirby. Harrington found an elegant reformulation of Paris's statement; his reformulation is a strengthening of the finite Ramsey's theorem on homogeneous sets

[Paris-Harrington 77]. This was followed by many papers by various authors, among them McAloon, Kotlarski, Murawski.

Later Paris and his students (Kirby, Clote, Kaye, Dimitroscopulos and others) turned to the study of fragments of PA. We shall rely substantially on their work. The first four chapters of the book deal mainly with fragments containing at least induction for Σ_1 -formulas. At present let us only say that in such theories we may freely construct recursive functions using primitive recursion. The fifth chapter deals with bounded arithmetic. Parikh seems to have been the first to study bounded arithmetic [Parikh]. He suggested investigating induction for bounded formulas since they are easily decidable (e.g. in linear space). This was developed significantly by Paris, Wilkie and Paris's students. The relation to complexity theory has been known from the beginning of the investigation of bounded arithmetic. Buss's dissertation, which later appeared as a book [Buss 86, Bounded ar.], was a further important impulse. Buss contributed both in finding new connections with complexity theory and in applying proof-theoretical methods. There are various later results; the reader will find such results here.

* * *

The aim of our study of the metamathematics of first-order arithmetic is to give the reader a deeper understanding of the role of the axiom schema of induction and of the phenomenon of incompleteness. In Part A, we develop important parts of mathematics and logic in various fragments of first order arithmetic. The main means are by coding of finite sets, arithmetization of logical syntax and semantics and through a version of König's lemma called the Low basis theorem.

Part B is devoted to incompleteness. Our main question reads: what more can we say about systems of arithmetic than that they are all incomplete? There are at least four directions in which the answer may be looked for:

(1) For each formula φ unprovable and non-refutable in an arithmetic T we may ask, how *conservative* it is over T , i.e. for which formulas ψ the provability of ψ in $(T + \varphi)$ implies the provability of ψ in T .

(2) We may further ask if $(T + \varphi)$ is *interpretable* in T , i.e. whether the notions of T may be redefined in T in such a way that for the new notions all axioms of $(T + \varphi)$ are provable in T .

(3) Given T we may look for *natural* sentences true but unprovable in T (for example, various combinatorial principles).

(4) Moreover, we may investigate *models* of T and look at how they visualize our syntactic notions and features.

Bounded arithmetic is studied in Part C. Various results of Part A are strengthened by showing that constructions done in stronger fragments are possible in some systems of bounded arithmetic and how. For bounded arithmetic we ask, besides questions (1)–(4), also the following:

(5) What is the relationship between provability in fragments and complexity of computation? One of the most important goals (presently inaccessible) is to show independence of some open problems of complexity theory from some fragments.

Details on the structure of the book are apparent from the table of contents.

Preliminaries

In this preliminary section, we first survey some basic facts from logic (and recursion theory) that are assumed to be known to the reader. Furthermore, we shall introduce the language of first order arithmetic and investigate first order definable sets of natural numbers. Finally, we shall present the beginnings of arithmetization of metamathematics by showing (or announcing) that various syntactic and some semantic logical notions can be understood as first order definable sets of natural numbers. To show that metamathematically interesting sets (like the set of all formulas, proofs, etc.) are (or can be understood as) first order definable sets of natural numbers is only the first step; the second step, more important and postponed until Chap. I, consists in investigating which first-order properties of these sets are provable in various systems of first order arithmetic. The fact that arithmetic can express its own syntax and partially its own semantics is of basic importance for the investigation of its metamathematics.

(a) Some Logic

0.1. Throughout the book, N is the set of all natural numbers (including zero). We shall denote natural numbers mainly by letters m, n, k, l , possibly indexed. The *least number principle* assures that each non-empty set of natural numbers has a least element. The *induction principle* says that if X is a set such that $0 \in X$ and X contains with each natural number n also its successor $n + 1$, then $N \subseteq X$.

0.2. Our survey of logic will have a double purpose: on the one hand, we shall investigate axiomatic systems of arithmetic as first-order theories and therefore first order logic will be our main device, and, on the other hand, we shall develop our axiomatic systems as meaningful mathematical theories and shall, among other things, formalize parts of first order logic in these systems. The fact that reasonable parts of logic can be developed in first-order arithmetic is of basic importance, as we shall see in the future.

0.3. A first-order *language* consists of *predicates* and *function symbols* (each predicate and function symbol has its non-zero natural *arity*), *constants*, and *variables*. A particular predicate = of *equality* (binary, i.e. of arity 2) is assumed to belong to each language. There are infinitely many variables. Constants and variables are *atomic terms*; if F is a k -ary function symbol and $t_0, \dots, t_k$ are terms then $F(t_0, \dots, t_k)$ is a *term*. An *atomic formula* is $P(t_0, \dots, t_k)$ where P is a k -ary predicate and $t_0, \dots, t_k$ are terms. If φ, ψ are formulas and x is a variable, then $\neg\varphi, \varphi \rightarrow \psi, (\forall x)\varphi$ are formulas. The symbols $\neg, \rightarrow$ are *connectives* (negation and implication); other usual connectives ($\&, \vee, \equiv$, etc.) are understood as abbreviations. $\forall$ is the *universal quantifier*; the *existential quantifier* $\exists$ is understood as an abbreviation. The notion of a *free* and *bound* variable in a formula is assumed to be known; e.g. x is free and y is bound in $P(x) \rightarrow (\forall y)Q(x, y)$. $\text{Subst}(\varphi, x, t)$ denotes the result of substitution of the term t for all free occurrences of the variable x in the formula φ . We often write $\varphi(x)$ instead of φ and $\varphi(t)$ instead of $\text{Subst}(\varphi, x, t)$ if there is no danger of misunderstanding.

0.4. A *model* for a language L consists of a non-empty domain M together with the following: for each k -ary predicate P of L , a k -ary relation $P_M \subseteq M^k$, for each k -ary function symbol F of L , a k -ary mapping $F_M : M^k \rightarrow M$, for each constant c an element $c_M \in M$. We use the same symbol M to denote both the model and its domain if there is no danger of misunderstanding. M has *absolute equality* if the equality predicate is interpreted by the identity relation $\{\langle a, a \rangle \mid a \in M\}$. An *evaluation* of a term in M is a finite mapping e whose domain consists of some variables, among them all variables occurring in t , and whose range is included in M . Similarly for an evaluation of a formula ($\text{dom}(e)$ contains all variables free in φ).

0.5. The *value* of a term t in a model M given by an evaluation e is defined as follows:

$$\begin{aligned} t_M[e] &\text{ is } t_M \text{ if } t \text{ is a constant,} \\ &e(t) \text{ if } t \text{ is a variable,} \\ &F_M(t_{1M}[e], \dots, t_{kM}[e]) \text{ if } t \text{ is } F(t_1, \dots, t_k). \end{aligned}$$

0.6. The following are *Tarski's conditions for satisfaction* ($M \models \varphi[e]$ is to be read " e satisfies φ in M ").

- (i) If φ is atomic, say $P(t_1, \dots, t_k)$, then $M \models \varphi[e]$ if $\langle t_{1M}[e], \dots, t_{kM}[e] \rangle \in P_M$ (the tuple of values of $t_1, \dots, t_k$ is in the relation that is the meaning of P).
- (ii) $M \models \neg\varphi[e]$ iff $M \not\models \varphi[e]$;
- (iii) $M \models (\varphi \rightarrow \psi)[e]$ iff $M \not\models \varphi[e]$ or $M \models \psi[e]$;
- (iv) $M \models (\forall x)\varphi[e]$ iff $M \models \varphi[e']$ for each e' coinciding with e on all arguments except x and defined for x .

0.7. Let Γ be a class of formulas of a language L , assume that Γ contains with each formula all its subformulas, let M be a model for L . A ternary relation *Sat* is a *satisfaction relation* for Γ in M if the following conditions hold:

- (1) *Sat* consists of some pairs $\langle \varphi, e \rangle$, where $\varphi \in \Gamma$ and e is an evaluation of φ .
- (2) Let $M \models \varphi[e]$ mean $\langle \varphi, e \rangle \in \text{Sat}$; then, for each $\varphi \in \Gamma$ and each evaluation e of φ , Tarski's conditions (i)–(iv) hold.

(Clearly, for each Γ and M , the satisfaction relation *Sat* for Γ in M is uniquely determined. But this is a rather strong fact; we shall investigate the provability of existence of various satisfaction classes in various axiomatic systems.)

0.8. φ is *true* in M ($M \models \varphi$) iff $M \models \varphi[e]$ for each e . We shall use various usual conventions in using the symbol $\models$; for example, if φ has the only free variable x and $a \in M$, we shall write $M \models \varphi[a]$ or $M \models \varphi(a)$ instead of $M \models \varphi[e]$ where e is the mapping defined only for x and giving x the value a .

0.9. A set $X \subseteq M$ is Γ -*definable* in M (where M is a model for L and Γ is a class of L -formulas) if there is a $\varphi \in \Gamma$ having exactly one free variable, such that $X = \{a \in M \mid M \models \varphi(a)\}$. (This is non-parametrical definability; we shall deal with parametrical definability later on.) Occasionally, we shall denote by φ_M the set defined by φ , thus: $a \in \varphi_M$ iff $M \models \varphi(a)$.

0.10. We shall fix any usual set of (Hilbert-style) *logical axioms* and *deduction rules*, for example the following ones:

Axioms:

$$\begin{aligned} & \varphi \rightarrow (\psi \rightarrow \varphi) \\ & ((\varphi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \chi))) \\ & (\neg\psi \rightarrow \neg\varphi) \rightarrow (\varphi \rightarrow \psi) \\ & (\forall x)\varphi(x) \rightarrow \varphi(t) \quad (t \text{ free for } x \text{ in } \varphi) \end{aligned}$$

Rules: From φ and $\varphi \rightarrow \psi$ infer ψ (modus ponens).

From $\nu \rightarrow \varphi(x)$ infer $\nu \rightarrow (\forall x)\varphi(x)$ if x is not free in ν .

0.11. An *axiomatic theory* in a language L is given by a set T of L -formulas called *special axioms* of the theory. Axioms for *equality* (saying that equality is reflexive, symmetric, transitive and is a congruence with respect to all predicates and function symbols) are assumed to belong to special axioms of each axiomatic theory; they will not be explicitly mentioned. $T \vdash \varphi$ means that φ is *provable* in T , i.e. there is a T -proof of φ (a sequence $\varphi_0, \dots, \varphi_n$ of L -formulas such that φ_n is φ and for each $i \leq n$, either φ_i is an axiom (logical

or special) or φ_i follows from some preceding members of the sequence using a rule of inference).

T is *consistent* if it does not prove any contradiction, i.e. for each φ , $T \not\vdash \varphi$ or $T \not\vdash \neg\varphi$ (or both).

M is a *model* of T if M is a model for the language L and each special axiom of T is true in M .

0.12 Gödel's Completeness Theorem. $T \vdash \varphi$ is true in each model of T iff φ is true in each countable model of T . Thus: T is consistent iff T has a model.

Convention. All models investigated in this book are countable (or finite).

Next we shall deal with Skolemizations. The reader is assumed to know how to convert each formula in a logically equivalent formula in the *prenex normal form*, i.e. a formula consisting of a block of quantifiers followed by an open (quantifier-free) formula.

0.13. Let T be a theory in a language L , let $\varphi(x_1, \dots, x_k, y)$ be an L -formula and let F be a k -ary function symbol not in L ; put $L' = L \cup \{F\}$. The formula

$$\varphi(x_1, \dots, x_k, y) \rightarrow \varphi(x_1, \dots, x_k, F(x_1, \dots, x_k))$$

is the *Skolem axiom* for φ and y .

0.14 Lemma. If T is a theory in a language L and $\hat{T}$ results from T by adding a Skolem axiom, then $\hat{T}$ is a conservative extension of T , i.e. each L -formula provable in $\hat{T}$ is provable in T .

(A model-theoretic proof is trivial: each (countable) model of T has an expansion to a model of $\hat{T}$. Indeed, let $M \models T$, and assume that the domain of M is N . For each $a \in N$, let $f(a)$ be the least $b \in N$ such that $M \models \varphi(a, b)$, if such a b exists; otherwise put $f(a) = 0$. Clearly, $(M, f) \models \hat{T}$.)

0.15. Let Φ be the formula

$$(Q_1 x_1), \dots, (Q_k x_k) \varphi(\mathbf{x}, \mathbf{y})$$

where Q_i is $\forall$ or $\exists$ ($= 1, \dots, k$). Let $\mathbf{x}$ mean $x_1, \dots, x_k$; let $\leftarrow x_i$ mean $x_1, \dots, x_i$; let $x_i \rightarrow$ mean $x_i, \dots, x_k$. Define a sequence of terms as follows:

$$\begin{aligned} t_i &= x_i \text{ if } Q_i \text{ is } \forall, \\ t_i &= F_i^{\Phi}(\leftarrow t_{i-1}) \text{ if } Q_i \text{ is } \exists, \end{aligned}$$

F_i^{Φ} being a new function symbol. Finally, put

$$sk(\Phi) = \varphi(t_1, \dots, t_k, \mathbf{y}).$$

(Example: $sk(\forall x)(\exists y)(\forall u)(\exists v)\varphi(x, y, u, v)$ is $\varphi(x, F_1(x), u, F_2(x, F_1(x), u))$.) If T is a theory, then $sk(T) = \{sk(\Phi) | \Phi \in T\}$.

0.16 Corollary. $sk(T)$ is an open conservative extension of T , i.e. all axioms of $sk(T)$ are quantifier-free and each L -formula provable in $sk(T)$ is provable in T .

Proof. For $i = 0, \dots, k$ let $\Phi^{(i)}$ result from Φ by deleting the first i quantifiers, thus $\Phi^{(i)}$ is $(Q_{i+1}x_{i+1}) \dots \varphi(x, y)$. First extend T by adding, for $i = 0, \dots, k$, the following Skolem axioms:

$$\Phi^{(i)}(\leftarrow x, y) \rightarrow \Phi^{(i)}(\leftarrow x_{i-1}, F_i^{\Phi}(\leftarrow x_{i-1}), y).$$

Do this for each axiom Φ of T . The new theory T' is a conservative extension of T . $\square$

Claim 1. $T' \vdash sk(T)$.

Take a $\Phi \in T$ and prove by induction $\Phi^{(i)}(\leftarrow t_i, y)$ in T' . $\Phi^{(i)}$ is Φ ; and $T', \Phi^{(i)}(\leftarrow t_i, y) \vdash \Phi^{(i+1)}(\leftarrow t_{i+1}, y)$ either by predicate calculus (if Q_{i+1} is $\forall$) or by the above Skolem axiom (if Q_{i+1} is $\exists$). And obviously $\Phi^{(k)}(t_1, \dots, t_k)$ is $sk(\Phi)$.

Claim 2. $sk(T) \vdash T$.

Prove by induction $sk(\Phi) \vdash \Phi^{(i)}(\leftarrow t_i, y)$ for $i = k, \dots, 0$. $\Phi^{(k)}(\leftarrow t_i, y)$ is $sk(\Phi)$; and $\Phi^{(i+1)}(\leftarrow t_{i+1}, y) \vdash \Phi^{(i)}(\leftarrow t_i, y)$ either by generalization (if Q_i is $\forall$) or by the logical schema $\alpha(t) \vdash (\exists x)\alpha(x)$ (if Q_i is $\exists$).

0.17 Lemma. Each theory T has an open conservative extension $\hat{T}$ in which each formula is equivalent to an open formula.

Proof. Put $T_0 = T$, T_{n+1} is the extension of T_n by Skolem axioms for all open formulas of T_n , let $T_\infty = \bigcup_n T_n$ and $T' = T_\infty - T_0$. Clearly, T_∞ is a conservative extension of T . We shall show that each formula ψ of T' is equivalent in T' to an open formula. For this purpose it suffices to assume ψ to have the form $(\exists y)\varphi(\mathbf{x}, y)$, φ open. But then the Skolem axiom for φ and y guarantees that, for an appropriate F , $T' \vdash (\exists y)\varphi(\mathbf{x}, y) \equiv \varphi(\mathbf{x}, F(\mathbf{x}))$. Now it suffices to replace in T_∞ each element of T_0 by its open equivalent; the resulting theory is $\hat{T}$. $\square$

0.18. For any Φ , let the *Herbrand variant* of Φ , $He(\Phi)$ be the existential closure of $\neg sk(\neg \Phi)$: e.g. if Φ is $(\forall x)(\exists y)(\forall u)(\exists v)\varphi(x, y, u, v)$, then $He(\Phi)$ is $(\exists y)(\exists v)\varphi(c, y, F(c, y, v))$.

0.19 Theorem. Φ is provable (in logic, i.e. in the theory with no special axiom) iff $He(\Phi)$ is provable.

(Immediate from 0.16.)

0.20 Lemma. Let $\varphi(\mathbf{x})$ be an open L -formula ($\mathbf{x}$ is a tuple of variables). The formula $(\exists \mathbf{x})\varphi(\mathbf{x})$ is provable (in logic) iff there are tuples $\mathbf{t}_1, \dots, \mathbf{t}_n$ of L -terms such that the disjunction

$$\varphi(\mathbf{t}_1) \vee \dots \vee \varphi(\mathbf{t}_n)$$

is a propositional tautology. (Each $\varphi(\mathbf{t}_i)$ is called an *instance* of $\varphi(\mathbf{x})$.)

Note that this also has an easy model-theoretic proof using Königs lemma; Königs's lemma will be studied in Chap. I, Sect. 3.

0.21 Herbrand's Theorem. A formula Φ is provable in logic iff there is a disjunction D of finitely many instances of the quantifier-free matrix of $He(\Phi)$ such that D is a propositional tautology.

This follows from the preceding. An elementary proof (not using model theory) can be found in Shoenfield's book. In I.4.15 we shall claim that Herbrand's theorem is (meaningful and) provable in a theory called IS_1 (defined in Chap. I, Sect. 1), again with the help of Shoenfield's book, and in III.3.30 we shall prove in IS_1 a theorem that has the implication $\Leftarrow$ of Herbrand's theorem as its corollary. (In fact, we shall elaborate Shoenfield's proof of that implication.) Finally, in Chap. V we prove Herbrand's theorem in a rather weak system of arithmetic.

We now turn to some basic notions and facts of recursion theory. Recall that N denotes the set of natural numbers.

0.22. Primitive recursive functions and general recursive functions are usually defined as follows:

$$\begin{aligned} \text{Basic PRF's:} \quad & Zero(n) = 0, \quad Succ(n) = n + 1, \\ & I_m^i(n_0, \dots, n_m) = n_i \quad (\text{where } 0 \leq i \leq m). \end{aligned}$$

A function $F : N^n \rightarrow N$ results from $G : N^m \rightarrow N$ and $H_1, \dots, H_m : N^n \rightarrow N$ by *composition* if

$$F(k_1, \dots, k_n) = G(H_1(k_1, \dots, k_n), \dots, H_m(k_1, \dots, k_n))$$

for each $k_1, \dots, k_n \in N$. An $F : N^{n+1} \rightarrow N$ results from $G : N_n \rightarrow N$ and $H : N_{n+2} \rightarrow N$ by *primitive recursion* if, for each $\mathbf{k} = (k_1, \dots, k_n)$, and each m ,

$$\begin{aligned} F(0, \mathbf{k}) &= G(\mathbf{k}), \\ F(m+1, \mathbf{k}) &= H(m, \mathbf{k}). \end{aligned}$$

The class of all *primitive recursive functions* (PRF's) is the smallest class containing basic PRF's and closed under composition and primitive recursion.

An $F : N_{m+1} \rightarrow N$ results from $G : N^{m+2} \rightarrow N$ by *regular minimization* if for each $m, \mathbf{k} = (k_1, \dots, k_n)$,

$$F(m, \mathbf{k}) = (\min q)(G(m, \mathbf{k}, q) = 0)$$

and for each $m, \mathbf{k}$ there exists a q such that $G(m, \mathbf{k}, q) = 0$ (so that F is total, i.e. defined for each $m, \mathbf{k}$).

The class of all *general recursive functions* is the smallest class containing the basic PRF's and closed under composition, primitive recursion and minimization.

0.23 Examples of PRF's: addition *Add*, multiplication *Mult*, exponentiation *Exp*, factorial *Fact*, difference *Diff*. We freely write $n + m, n * m, n!, n - m$ instead of $Add(n, m), Mult(n, m), Exp(n, m), Fact(n), Diff(n, m)$, respectively. (A word on difference: $n - m$ for natural numbers means $\max(n - m, 0)$ as meaningful for integers; thus $5 - 3 = 2$ and $3 - 5 = 0$.)

0.24. A set $X \subseteq N_n$ is *primitive recursive* (PR) [*general recursive* (GR)] if its characteristic function

$$\chi_X(k_1, \dots, k_n) = \begin{cases} 1 & \text{if } \langle k_1, \dots, k_n \rangle \in X, \\ 0 & \text{otherwise.} \end{cases}$$

is PR [GR, respectively].

0.25 Examples. The equality relation as well as the less-than relation are both primitive recursive; both PR and GR sets are closed under Boolean operations. The set of all primes is a PR set; the increasing enumeration p_n of primes ($p_0 = 2, p_1 = 3, p_2 = 5, p_3 = 7, p_4 = 11$ etc.) is a PRF.

0.26. Let Γ be a class of functions such that each $F \in \Gamma, F : N^n \rightarrow N$ for some n . (We say that Γ is a class of *total number theoretic functions*. It is obvious what we mean by saying that Γ is closed under substitution, primitive recursion, regular minimization, etc. A Γ set (relation) is a set (relation) whose characteristic function is in Γ . If Γ contains basic PRF's and is closed under composition and primitive recursion (or: under composition and regular minimization) then it is closed under definitions of functions by cases (with a condition in Γ) and under bounded minimization. In more detail:

Let A be a Γ set, let $F_1, F_2 : N \rightarrow N$ be in Γ . Define

$$\begin{aligned} F(n) &= F_1(n) \text{ if } n \in A, \\ F(n) &= F_2(n) \text{ otherwise.} \end{aligned}$$

Then $F \in \Gamma$. (Generalize for $F_1, \dots, F_k$ of n arguments and $A_1, \dots, A_k$ a partition of N^n .)

Let $R \subseteq N^{n+1}$, let R be a Γ -relation and put

$$\begin{aligned} F(k, \mathbf{q}) &= (\min m \leq k) R(m, \mathbf{q}) \text{ if there is such an } m, \\ F(k, \mathbf{q}) &= 0 \text{ otherwise;} \\ S(k, \mathbf{q}) &= \{ \langle k, \mathbf{q} \rangle \mid (\exists m \leq k) R(m, \mathbf{q}) \}. \end{aligned}$$

Then $F \in \Gamma$ and S is a Γ -relation.

0.27. For each class Γ of number-theoretic total functions, let $\text{Prim}(\Gamma)$ (the class of all functions primitive recursive in Γ) be the minimal class containing all basic primitive recursive functions, all elements of Γ and closed under composition and primitive recursion. Similarly for the class $\text{Rec}(\Gamma)$ of all functions general recursive in Γ .

(b) The Language of Arithmetic, the Standard Model

0.28. Recall that N is the set of natural numbers. N also denotes the set of natural numbers together with the usual arithmetical structure:

- the unary operation *Succ* of successor (adding one),
- the binary operation *Add* of addition,
- the binary operation *Mult* of multiplication,
- the binary relation *Ord* of linear order,
- the minimal element 0.

N is certainly a very natural and very mathematical structure, the ground stone of mathematics. We introduce a first order language L_0 such that N is a model of this language. L_0 has

- a unary function symbol S ,
- binary function symbols $+$, $*$,
- the equality predicate $=$,
- a binary predicate $\leq$,
- a constant 0.

L_0 is the language of first-order arithmetic and N is its standard model. Note that each natural number n is named by a variable-free term $\bar{n}$ of L_0 : we can just take $\bar{n}$ to be $S(S(\dots S(0)\dots))$ (n occurrences of S). Thus 1 is $S(0)$, 4 is $S(S(S(S(0))))$, etc. For some investigations (in Chap. V) we need more economical names; this will be made explicit if the situation demands. The term $\bar{n}$ is the n th numeral.

Notational Conventions. We shall freely use obvious conventions in writing terms of L_0 : first, we shall use the infix notation (we write $x + y$ rather

than $+(x, y)$, the same for $*$), *second*, the multiplication sign may be omitted if there is no danger of misunderstanding (xy means $x * y$), *third*, we omit unnecessary parentheses, declaring $*$ to be superordinated to $+$ ($x * y + 2$ and $xy + 2$ both stand for $(x * y) + 2$ etc.).

0.29. Any model isomorphic to N is also called standard. It is easy to show that there is a model M which is elementarily equivalent to N (i.e. has the same true L_0 -formulas) but is not standard: let $Th(N)$ be the set of all sentences true in N , let c be a new constant and let $T = Th(N) \cup \{\bar{n}(c) \mid n \in N\}$. By compactness, T is consistent and hence has a model M . Show by induction that if f is an isomorphism of N to M then for each n , $f(n) = \bar{n}_M$ and therefore c_M has no preimage. Thus M is not isomorphic to N .

0.30 Bounded Quantifiers and Arithmetical Hierarchy. $(\exists x \leq y)\varphi$ is an abbreviation for $(\exists x)(x \leq y \ \& \ \varphi)$ and $(\forall x \leq y)$ is an abbreviation for $(\forall x)(x \leq y \rightarrow \varphi)$. By convention, x and y must be *distinct variables*. An L_0 -formula is *bounded* if all quantifiers occurring in it are bounded, i.e. occur in a context as above. Furthermore, $(\forall x < y)\varphi$ is an abbreviation for $(\forall x \leq y)(x \neq y \rightarrow \varphi)$ and similarly for $(\forall x < y); x \neq y$ is the same as $\neg(x = y)$.

We introduce a hierarchy of formulas called the *arithmetical hierarchy*. Σ_0 -formulas = Π_0 -formulas = bounded formulas; Σ_{n+1} -formulas have the form $(\exists x)\varphi$ where φ is Π_n , Π_{n+1} -formulas have the form $(\forall x)\varphi$ where φ is Σ_n . Thus a Σ_n -formula has a block of n alternating quantifiers, the first one being existential, and this block is followed by a bounded formula. Similarly for Π_n .

0.31. A set $X \subseteq N$ is Σ_n (or Π_n) if it is defined by a Σ_n -formula ($(\Pi_n$ -formula) with exactly one free variable. Similarly for a relation $R \subseteq N^k$. X is Δ_n if it is both Σ_n and Π_n . A function $F : N^k \rightarrow N$ is Σ_n , etc., if it is Σ_n as a relation $\subseteq N^{k+1}$ (the graph of F).

In particular, X is Δ_0 iff it is Σ_0 ; Π_n relations are complements of Σ_n relations and vice versa.

0.32 Pairing. There is a Σ_0 pairing function, i.e. a one-one mapping OP of N_2 onto N , increasing in both arguments.

Indeed, the usual "diagonal" enumeration of ordered pairs of natural numbers

	0	1	2	3	...
0	0	1	3	6	...
1	2	4	7	...	
2	5	8	...		
3	9	...			

satisfies the following:

$$OP(m, n) = \frac{1}{2}(m + n + 1)(m + n) + m.$$

Clearly, this function is defined by the formula

$$2z = (x + y + 1)(x + y) + 2x;$$

we denote the last formula by $OP(x, y, z)$. Furthermore, we expand N by adding OP to its structure; and expand L_0 by a new binary function symbol (x, y) interpreted as OP . We keep the notation N, L_0 for the (inessentially) expanded structure and language. Thus we have

$$N \models (\forall x, y) OP(x, y, (x, y))$$

and for each $m, n \in N$ we have

$$OP(m, n) = (m, n)_N$$

If there is no danger of misunderstanding we omit the subscript N in $(m, n)_N$; thus we write also (m, n) for $OP(m, n)$.

0.33 Notation Conventions Continued. We give a detailed notational explanation on the pairing function since this exemplifies a general notational method common in the metamathematics of arithmetic and also used in the present book:

- (1) The structure N and language L_0 is notationally not distinguished from its inessential expansions if not necessary.
- (2) If we have a relation $R \subseteq N^k$ and exhibit a concrete definition of R in N formulated in L_0 then the defining formula is denoted by $R^\bullet$ (dot notation). Similarly for functions.
- (3) Conversely, if we have a function symbol F and its interpretation F_N in N we often omit the subscript N and write $F(k, \dots)$ instead of $F_N(k, \dots)$. Similarly for relations.

Now that we have introduced the language of arithmetic we see that $m + n$ is shorthand for $m +_N n$ and that the formula $x + y = z$ could be denoted by $Add^\bullet$; similarly for $Succ$ and $Mult$.

This convention will be used tacitly through the book; it will be generalized (and made more precise) in connection with axiomatic theories having N as one of their models.

Caution. Even if we expand the language we keep the notion of Σ_n and Π_n formulas unchanged, i.e. assume that they are formulated in L_0 in its original meaning. (A formula in the enriched language may or may not be *equivalent* to a Σ_n or Π_n formula; this needs further investigation).

0.34 Theorem. For each natural n ,

- (1) Σ_n , Π_n , Δ_n relations are closed under intersection and union;
- (2) Δ_n relations are closed under complementation;
- (3) if $n > 0$ then Σ_n relations are closed under existential projection and Π_n relations are closed under universal projection.

Proof. We prove (1) & (2) & (3) by induction on n . For $n = 0$ the assertion is evident. Assume it for n and consider $n + 1$. The claim (2) is trivial; let us prove (3) for Σ_{n+1} (the proof for Π_{n+1} is similar). Let R be defined by $(\exists z)\varphi(\mathbf{x}, y, z)$ where φ is Π_n , and let R' be defined by $(\exists y)(\exists z)\varphi(\mathbf{x}, y, z)$. Then R' is defined by

$$(\exists u)(\forall y)(\forall z)(u = (y, z) \rightarrow \varphi(\mathbf{x}, y, z))$$

as well as by

$$(\exists u)(\forall y \leq u)(\forall z \leq u)(u = (y, z) \rightarrow \varphi(\mathbf{x}, y, z)) .$$

If $n = 0$ then the latter formula is clearly Σ_1 ; if $n > 0$ then, by the induction assumption, the former formula is equivalent (in N) to a Σ_{n+1} formula. (Once and for all, let us elaborate details: φ is Π_n , both $u = (y, z)$ and its negation are Σ_0 , hence Π_n , and by (3), the formula in question is also Π_n .)

To prove (1) let $(\exists y)\varphi(\mathbf{x}, y)$ and $(\exists z)\psi(\mathbf{x}, z)$ be Σ_{n+1} and assume y, z to be distinct variables. Then $(\exists y)\varphi(\mathbf{x}, y) \& (\exists z)\psi(\mathbf{x}, z)$ is logically equivalent to $(\exists y)(\exists z)(\varphi(\mathbf{x}, y) \& \psi(\mathbf{x}, z))$ and similarly for $\forall$; thus (1) for n and (3) for $(n + 1)$ give the result. $\square$

0.35 Theorem. Each Σ_0 set is primitive recursive.

Proof. Since successor, addition and multiplication are PRF's, each term defines a PRF; since equality and ordering are PR relations, each atomic formula defines a PR relation. Dummy variables may be introduced using I_m^i . And PR relations are closed under Boolean operations and bounded projection. $\square$

We shall now investigate the question whether each PRF, and moreover, each GRF, is definable in N . The result will be that general recursive functions coincide with Δ_1 functions; this appears to show that the choice of our language is natural. First note the following

0.36 Lemma. If a function $F : N^n \rightarrow N$ is Σ_1 then it is Δ_1 .

Proof. Let F be defined by a Σ_1 formula $\varphi(\mathbf{x}, y)$, i.e. $F(m_1, \dots) = k$ iff $N \models \varphi(m_1, \dots, k)$. Then the complement of F in N^{n+1} is defined by $(\exists z)(z \neq$

$y \& \varphi(x, z)$ which is again a Σ_1 formula. Note that the lemma does not generalize to partial functions, i.e. mappings from N^n into N . $\square$

0.37 Lemma. Basic PRF's are defined by open formulas.

Proof. Take $y = 0$, $y = S(x)$, $y = x_i$. $\square$

0.38 Lemma. Δ_1 functions are closed under composition.

Proof. For simplicity, let $F(k) = G(H(k))$ for each k , and let $\varphi(x, y), \psi(x, y)$ define G, H respectively, $\varphi, \psi \in \Sigma_1$. Then F is defined by the Σ formula

$$(\exists z)(\psi(x, z) \& \varphi(z, y)) .$$

$\square$

0.39 Lemma. Σ_1 relations are closed under bounded universal projections.

Proof. Let $R \subseteq N^2$ be defined by a formula $(\exists z)\varphi(x, y, z)$ where φ is Σ_0 and let $S \subseteq N$ be defined by $(\forall x \leq y)(\exists z)\varphi(x, y, z)$. We show that S is also defined by $(\exists w)(\forall x \leq y)(\exists z \leq w)(\varphi(x, y, z))$, which is Σ_1 . (Thus the quantifier $(\exists z)$ can be bounded.) Clearly the latter formula implies the former. Thus assume $k \in S$; we find a q such that $N \models (\forall x \leq \bar{k})(\exists z \leq \bar{q})\varphi(x, \bar{k}, z)$. To this end we show by induction that for each $i = 0, 1, \dots, k$ there is a q_i such that

$$N \models (\forall x \leq \bar{i})(\exists z \leq \bar{q}_i)\varphi(x, k, z) .$$

Since $k \in S$ we know $N \models (\forall x \leq k)(\exists z)\varphi(x, k, z)$; thus the case $i = 0$ is evident. Assume q_i has been found and let r be such that $N \models \varphi(\overline{i+1}, \bar{k}, r)$. Put $q_{i+1} = \max(q_i, r)$. $\square$

0.40 Lemma. Δ_1 functions are closed under regular minimization.

Proof. Let $F(k) = (\min q)(G(k, q) = 0)$, $F : N \rightarrow N$, G be Σ_1 defined by $\varphi((x, y, z))$. Then F is Σ_1 defined by

$$\varphi(x, y, 0) \& (\forall y' < y)(\exists z \neq \bar{0})\varphi(x, y', z) .$$

This shows that F is Σ_1 , hence, by 0.36, it is Δ_1 . $\square$

The problem is to show that Δ_1 functions are closed under primitive recursion. If F results from G from G and H by primitive recursions then an explicit definition of $F(k)$ is easily made using the sequence $F(0), F(1), \dots, F(k)$ since we can describe $F(0)$ and describe $F(i+1)$ from $F(i)$. Thus some Δ_1 definable coding of finite sequences of natural numbers by natural numbers

is desirable. In fact, such a coding is a device used very often in arithmetic. We shall state the existence of such a coding using the following

0.41 Definition. A coding of finite sequences (of natural numbers by natural numbers) consists of a PR set $Seq \subseteq N$ and PRF's

- lh (unary; $lh(s)$ is called the length of s),
- $memb$ (binary; $memb(s, i)$ is the i th member of s),
- $prolong$ (binary; $prolong(s, k)$ is the result of juxtaposing k with s)

such that the following holds for each $s, s' \in Seq$:

- (1) $lh(s) \leq s$ and, for each $i < lh(x)$, $memb(s, i) < s$;
- (2) there is an empty sequence $\emptyset$ with $lh(\emptyset) = 0$;
- (3) for each $k \in N$ if $s' = prolong(s, k)$ then $lh(s') = lh(s) + 1$, for $i < lh(s)$ we have $memb(s, i) = memb(s', i)$ and for $i = lh(s)$ we have $memb(s', i) = k$.
- (4) (monotonicity): if $lh(s) \leq lh(s')$ and, for each $i < lh(s)$, $memb(s, i) \leq memb(s', i)$ then $s \leq s'$;
- (5) the set $N - Seq$ is infinite.

(Note that (4) implies extensionality; if s, s' have the same length and the same corresponding members then they are equal.)

0.42 Theorem. There is a Δ_1 coding of finite sequences; i.e. a coding such that the set Seq and the functions $lh, memb, prolong$ are Δ_1 (besides being PR).

The proof of this theorem is put off until Chap. I, Sect. 1; we shall then show more, namely that the properties of the coding are *provable* in a suitable fragment of arithmetic.

For most investigations of Chaps. I-IV it is immaterial which concrete coding of sequences is taken; but for some more subtle results, especially on weak fragments, special care will be necessary. In fact, we prove in Chap. V that there is a Σ_0 coding of finite sequences.

Notation. The chosen Δ_1 definitions of $Seq, lh, memb$ and $prolong$ will be denoted by $Seq^\bullet, lh^\bullet, memb^\bullet$ and $prolong^\bullet$; $lh^\bullet$ and $prolong^\bullet$ will also be used as function symbols, thus we shall write $y = lh^\bullet(x)$ instead of $lh^\bullet(x, y)$.

We expand L_0 by a new function (symbol $(-)_y$ for the y -th member of x (thus in formulas we write $z = (x)_y$ for $memb^\bullet(x, y, z)$).

And if there is no danger of misunderstanding, we shall use this bracket notation also informally, thus $(s)_i$ will be the same number as $memb(s, i)$.

A similar convention for the function $prolong$ will be made later.

0.43 Corollary. Δ_1 functions are closed under primitive recursion.

Proof. Assume $F(0) = m$ and $F(k+1) = H(k, F(k))$; let H be defined by $\kappa(x, y, z)$. Then F is defined by the following formula $\varphi(x, y)$:

$$(\exists z)(Seq^*(z) \ \& \ lh^*(z) = x + 1 \ \& \ (z)_{\bar{0}} = \bar{m} \ \& \\ (\forall u < lh^*(z))(\forall v < u)(v + 1 = u \rightarrow \kappa(v, (z)_v, (z)_u))$$

Similarly for the case of F having parameters. □

0.44 Remark. (1) In particular, *exponentiation* ($n = m^k$) is Δ_1 since it is primitive recursive. We shall show in Chap. V that exponentiation is Δ_0 (which is a rather non-trivial result).

(2) An apparently more general form of primitive recursion defines $F(k+1)$ from the course of values $F(0), \dots, F(k)$ directly. Let, for each F , $\hat{F}(k, \mathbf{m}) = s$ iff s is the (code of the) sequence of length $k+1$ such that for each $i \leq k$, $(s)_i = F(i, \mathbf{m})$. F results from G, H by *primitive recursion on the course of values* if $F(0, \mathbf{m}) = G(\mathbf{m})$ and $F(k+1, \mathbf{m}) = H(k, \hat{F}(k), \mathbf{m})$. Clearly, Δ_1 functions are closed under this kind of primitive recursion.

(3) If the reader has a favourite primitive recursive coding of sequences he may keep it since now he knows that his coding is Δ_1 which is sufficient for most applications. But he should keep in mind that it might be rather difficult and cumbersome to show directly that his coding is Δ_1 (or even Σ_0).

0.45 Theorem. A function $F : N_n \rightarrow N$ is general recursive iff it is Δ_1 .

Proof. Clearly, each GRF is Δ_1 since basic functions are and the class of Δ_1 functions is sufficiently closed.

Conversely, if $F : N \rightarrow N$ is Δ_1 , thus $F(k) = n$ iff $N \models (\exists z)\varphi(\bar{k}, \bar{n}, z)$ where φ is Σ_0 then by 0.35, the relation $R \subseteq N^3$ defined by φ is primitive recursive. Define $F_0(k)$ to be the least sequence s of length 2 such that $R(k, (s)_0, (s)_1)$; then $F(k) = (F_0(k))_0$. F_0 results from F by a regular minimization and taking the 0-th member of a sequence is a primitive recursive function; thus F is a GRF. □

0.46 Fact. An infinite Δ_1 set $X \subseteq N$ has an infinite increasing enumeration (i.e. a $F : N \rightarrow N$ mapping N one-one and increasing onto X).

(The reader can either use the fact that this is true for recursive sets of natural members or prove that fact directly, which is easy using the available means.)

0.47 Some Useful PRFs Concerning Sequences.

(1) For each $n \geq 1$, there is an n -ary PRF associating with each $k_0, \dots, k_{n-1} \in N$ the n -tuple $\langle k_0, \dots, k_{n-1} \rangle$, i.e. the sequence s of length n such that, for each $i < n$, $(s)_i = k_i$.

(2) *Concatenation*: For $s, t \in Seq$, $s \frown t$ denotes the *concatenation* of s, t , i.e. the sequence w such that

$$\begin{aligned} lh(w) &= lh(s) + lh(t), \\ (w)_i &= (s)_i \text{ for each } i < lh(s), \\ (w)_{lh(s)+j} &= (t)_j \text{ for each } j < lh(t). \\ \text{Put } s \frown t &= 0 \text{ if } s \notin Seq \text{ or } t \notin Seq. \end{aligned}$$

We show that this function is primitive recursive.

$$\begin{aligned} \text{Define } C(s, t, 0) &= s \\ C(s, t, i+1) &= \text{prolong}(C(s, t, i), (t)_i) \text{ if } i < lh(t), \\ C(s, t, i+1) &= C(s, t, i) \text{ if } i \geq lh(t), \\ \text{put } s \frown t &= C(s, t, lh(t)). \end{aligned}$$

(3) *Concatenation of a sequence of sequences*. If $w \in Seq$ and for each $i < lh(w)$, $(w)_i \in Seq$ then put

$$Concseq(w) = (w)_0 \frown (w)_1 \frown \dots \frown (w)_{lh(w)-1}$$

Concseq is primitive recursive:

Define

$$\begin{aligned} D(w, 0) &= \emptyset, \\ D(w, i+1) &= D(w, i) \frown (w)_i \quad \text{if } i < lh(w), \\ D(w, i+1) &= D(w, i) \quad \text{if } i \geq lh(w), \\ Concseq(w) &= D(w, lh(w)). \end{aligned}$$

The reader may easily verify the following facts for sequences s, t ($s \subseteq t$ means that s is an *initial segment* of t , i.e. $lh(s) \leq lh(t)$ and for each $i < lh(s)$, $(s)_i = (t)_i$);

- (1) $s \frown t \subseteq s \frown t'$ implies $t \subseteq t'$,
- (2) $s \frown t \subseteq s' \frown t'$ implies $s \subseteq s'$ or $s' \subseteq s$,
- (3) $s \subseteq t$ implies the existence of a unique u such that $t = s \frown u$,
- (4) $Concseq(s \frown t) = Concseq(s) \frown Concseq(t)$.

0.48 Matiyasevič(-Robinson-Davis-Putnam) Theorem. Σ_1 relations coincide with relations defined by existential L_0 -formulas, i.e. formulas consisting of a block of existential quantifiers followed by an open formula.

We may additionally assume that the open formula in question does not contain the predicate $\leq$ (thus atomic formulas are only equalities of terms)

since $x \leq y$ may be replaced by $(\exists z)(z = x = y)$ and $\neg(x \leq y)$ by $y \leq x \ \& \ x \neq y$. Thus each open formula containing $\leq$ is equivalent to an existential formula not containing $\leq$.

A readable proof may be found in [Davis 73, Hilbert's tenth]. Note that this theorem (often called the MRDP theorem) is very famous; it implies recursive unsolvability of Hilbert's tenth problem.

0.49 Remark. Concerning the choice of the language L_0 , observe that what we have said till now gives some justification to our choice of the language of arithmetic. In this language, all GRF's are first order definable (which is very natural for a first order arithmetic); and it can be shown that multiplication is not first order definable in the reduct of N to $(L_0 \text{ without } *)$ and similarly, addition is not first order definable using $(L_0 \text{ without } +)$.

This follows from the fact that the set of all sentences of $(L \text{ without } *)$ true in N is Δ_1 (i.e. recursive), the same for sentences of $(L \text{ without } +)$ and from the undecidability results of Chap. III.

On the other hand, zero, successor and ordering are easily definable in the reduct of N to $(+, *)$; the reasons for taking them as primitives are only technical and inessential variants are possible.

(c) Beginning Arithmetization of Metamathematics

0.50 Introduction. To *arithmetize metamathematics* means to make metamathematics a part of arithmetic (or at least to make important *parts* of metamathematics parts of arithmetic). It is Gödel's invention that this is possible. The first task consists in showing that important logical notions are *definable* in N by formulas of first order arithmetic; this is our task in the present subsection. The second task is then to show that important properties of these notions are provable in various systems of axiomatic arithmetic. (This task is postponed.)

To be able to define logical notions by arithmetical formulas we must identify objects of logic (as symbols, formulas, proofs, etc.) with numbers. There are two approaches to this task, not substantially different. First, we may think of logical objects as non-numbers (whatever they may be) and give some explicit rules on how to associate numbers to them. This procedure is usually called Gödel numbering and speaks of Gödel numbers of formulas, proofs, etc. Feferman observed that we have another apparently simpler possibility: just to *identify* logical objects with some numbers.

Recall our (pseudo)definition of terms: we defined some atoms (atomic terms) and specified operations (formation rules) under which the set of terms is closed. There are two tacit assumptions: first that the set of terms is the *least* set containing all atoms and closed under formation rules; and, second, that each non-atom t uniquely determines the formation rule and

its components that give t according to the formation rule. Similarly for formulas; so let us speak generally about *expressions*. We have a set $At \neq \emptyset$ of *atoms*, a set Op of *operations*, each operation e having its *arity* $Ar(e)$, and expressions are just elements of the free algebra generated by our atoms using our operations. More precisely, the *free algebra* of the type (Op, Ar) generated by At is a set $Expr \subseteq At$ together with a function $Appl$ (of application) associating with each operation o , and each sequence s of expressions such that $lh(s) = Ar(o)$, an expression $Appl(o, s) \in At$ such that $Appl$ is one-one (for such pairs (o, s)) and $Expr$ is the smallest set containing At and closed under $Appl$. Generalizing slightly, we replace the assumption $At \subset Expr$ by the assumption that we have a one-one embedding of At into $Expr$; it will be technically convenient to assume that for each atom $a \in At$ the one-element sequence $\langle a \rangle$ is an *atomic expression*. $Appl$ is then defined for pairs (o, s) as above and its range is the set of non-atomic expressions.

Finally, two free algebras given by At, Op, Ar are isomorphic in the obvious sense. Thus we may speak of *the* free algebra and its various *presentations*. We are interested in Δ_1 presentations.

0.51 Fact. Let $\emptyset \neq At \subset N$, let (Op, Ar) be a type, $At \cap Op = \emptyset$. Then there is a presentation $(Expr, Appl)$ of the free algebra of the type (Op, Ar) generated by At such that both the set $Expr$ and the function $Appl$ are primitive recursive in (At, Op, Ar) .

Proof. For each $o \in Op$ and each sequence s of length $Ar(o)$ let $Appl(o, s)$ be $\langle o \rangle \frown Concseq(s)$, i.e. the sequence beginning by o and continuing by the concatenation of all members of s ; let $Appl(o, s) = 0$ otherwise. (Note that this presentation is often called the Polish notation.) Clearly, $Appl$ is PR in (Op, Ar) . Call w a *derivation* of z if w is a sequence, its last element is z and for each $i < lh(w)$ we have the following:

either $(w)_i$ is an atomic expression $\langle x \rangle$ or there are $o, s < w$ such that $(w)_i = \langle o \rangle \frown Concseq(s)$, $o \in Op$, s is a sequence of length $Ar(o)$ and for each $k < lh(s)$, there is a $j < i$ such that $(s)_k = (w)_j$ (i.e. $(w)_i$ results from some preceding elements of w using an operation).

Let

$$Expr = \{z | (\exists w)(w \text{ is a derivation of } z)\}.$$

We show that $(Expr, Appl)$ is a presentation of the free algebra in question. $\square$

Lemma A. If e, e' are expressions and $e \subset e'$ then $e = e$.

Proof. Let e be the smallest expression such that there is an expression e' which is a proper initial segment of e . Then $e = \langle o \rangle \frown Concseq(s)$ and $e' = \langle o \rangle \frown Concseq(s')$, $s \neq s'$. Let i be the least number such that $(s)_i \neq (s')_i$;

show (using 0.47 (1)–(4)) that $(s)_i \subset (s')_i$ or $(s')_i \subset (s)_i$ and $(s)_i, (s')_i$ are expressions less than e . $\square$

Lemma B. If $e = \langle o \rangle \frown \text{Concseq}(s)$ and $e' = \langle o \rangle \frown \text{Concseq}(s')$ are expressions and $e = e'$ then $s = s'$.

Proof. Assume not; then $\text{Concseq}(s) = \text{Concseq}(s')$ and if i is the least such that $(s)_i \neq (s')_i$ then $(s)_i \subset (s')_i$ or $(s')_i \subset (s)_i$, which contradicts Lemma A. Thus $(\text{Expr}, \text{Appl})$ is a presentation.

It remains to show that Expr is a set PR in $(\text{At}, \text{Op}, \text{Ar})$. For this it is sufficient to bound the quantifier $(\exists w)$ in the definition above, i.e. to find a function H PR in $(\text{At}, \text{Op}, \text{Ar})$ such that

$$\text{Expr} = \{e | (\exists w < H(e))(w \text{ is a derivation of } e)\}.$$

To this end show that if e has a derivation then it has a derivation w' without repetitions and such that each $(w)_i$ is a (non-initial) segment of e , i.e. for some s, t , $e = s \frown (w)_i \frown t$. (Just omit all superfluous members of w and show that the resulting sequence w' is a derivation of e).

We know from the preceding that for each s there is at most one expression e' and at most one t such that $e = s \frown e' \frown t$; thus sequence w' satisfies $lh(w') \leq lh(e)$. Thus we can choose $H(e) = (e, \dots, e)$ (e times); clearly, H is PR. This completes the proof of 0.51. $\square$

0.52 Corollary. If $(\text{At}, \text{Op}, \text{Ar})$ is PR then $(\text{Expr}, \text{Appl})$ is PR; if the former is Δ_1 then the latter is Δ_1 .

0.53 Definition. A first order language is Δ_1 if the sets of all predicates, function symbols, constants and variables are (mutually disjoint) Δ_1 sets and the function Ar defined for each predicate and function symbol (arity) is a Δ_1 function. We additionally assume that no predicate, function symbol, constant and variable is a sequence and that there are two further non-sequences denoted $\neg, \rightarrow$.

0.54 Corollary. If a language L is Δ_1 then there are Δ_1 sets Term (of all terms) and Form (of all formulas) such that

- (1) Term is the free algebra given by variables and constants as atoms and function symbols with their arities as operations;
- (2) The set of all atomic formulas is Δ_1 ; the functions associating with each atomic formula its predicate and its sequence of arguments respectively are Δ_1 ; and no atomic formula is a sequence.
- (3) Form is the free algebra given by atomic formulas as atoms and by the following operations: $\rightarrow$ (binary), $\neg$ (unary) and for each variable x an operation $(\forall x)$ (unary).

0.55 Discussion. Here we stop our preliminary development of arithmetization. We survey ideas that could follow; we shall not elaborate on them here since we shall prove stronger results in Chap. I that will imply the facts sketched below as corollaries. Namely, instead of showing that some things are Δ_1 definable in the standard model, i.e. that some definition have some properties in N we show that these properties are *provable* in some fragments of arithmetic. We shall prove in particular the following:

- the substitution function *Subst* is Δ_1 in N ;
- the set of all logical axioms is Δ_1 in N . A theory is *axiomatized* if its language is Δ_1 and its set of special axioms is also Δ_1 .

It is easy to see that for each axiomatized theory T the set of all *proofs in T* (T -proofs) is Δ_1 and the set of all T -provable formulas is Σ_1 . T is *decidable* if the set of T -provable formulas is Δ_1 . (Undecidability of axiomatized systems of arithmetic is closely related to their incompleteness and will be studied in Part B of the book.)

Concerning semantics:

- the evaluation function *Val* of terms in N is Δ_1 in N ;
- the satisfaction for Σ_0 formulas in N is Δ_1 in N .

In Chap. I we shall show that basic facts about arithmetization as sketched till now are provable in the theory $I\Sigma_1$ using induction for Σ_1 formulas. This will be basic for our investigations of systems of arithmetic containing $I\Sigma_1$, which are a matter of interest in the main part of the book. But note that Chap. V is devoted to theories weaker than $I\Sigma_1$; in these theories special care is necessary and special codings of sequences, formulas etc. are used. Chapters I–IV occasionally use some results from Chap. V; explicit reference will always be made.

How precious also are thy thoughts unto me,
O God! how great is the sum of them! If I
should count them, they are more in number
than the sand; when I awake, I am still with
thee.

(Psalm 139, 17–18)

Part A

Positive Results on Fragments of Arithmetic

Chapter I

Arithmetic as Number Theory, Set Theory and Logic

Introduction

We are going to investigate axiomatic theories formulated in the language L_0 of arithmetic. Such a theory T is *sound* if the standard model N is a model of T , i.e. all axioms of T are true in N . If T is sound then, trivially, each formula provable in T is true in N . We confine our attention to theories containing a rather weak finitely axiomatized theory Q (which will be defined in a moment) and shall study an infinite hierarchy of sound theories whose union is called *Peano arithmetic*; the theories from the hierarchy are called *fragments* of Peano arithmetic. In this chapter and the next, we shall elaborate *positive* results on these theories, i.e. we shall show that the expressive and deductive power of these fragments is rather big: our aim will be to show how some amount of arithmetization of metamathematics yields the possibility of speaking inside a fragment of arithmetic not only of numbers but also of finite sets and sequences and of definable infinite sets of numbers. This is the main result of Sect. 1. In Sect. 2 we shall study the structure of the hierarchy of fragments, i.e. show various equivalent axiomatizations and several inclusions among fragments. Section 3 is devoted to the development of some recursion theory in fragments, notably to a proof of the Low basis theorem, which can be viewed as a strong form of König's lemma. (The Low basis theorem will be crucial in proofs of combinatorial principles in fragments; this will be done in Chap. II.) Finally Sect. 4 further develops metamathematics in fragments; among other things, the Low arithmetized completeness theorem, i.e. a strong form of the completeness theorem, will be proved.

Let us close this introduction with two remarks: first, the reader will find here (in Part A) actual *proofs* of various theorems in fragments, not only proofs of provability of these theorems. (Model-theoretic methods of proving provability of a sentence in a fragment can be found in Chap. IV.) It is hoped that the reader will feel comfortable in these fragments and will gain good practice in proving theorems in them. If so, then he will agree that each

fragment (as well as the whole of Peano arithmetic) captures a natural part of the truth about N .

Secondly, the *limitations* of the axiomatic approach in capturing the truth on natural numbers, i.e. the feature of *incompleteness*, will be studied in Part B.

1. Basic Developments; Partial Truth Definitions

(a) Properties of Addition and Multiplication, Divisibility and Primes

1.1 Definition. Q is the theory in the language L_0 with the following axioms:

- | | |
|------|--|
| (Q1) | $S(x) \neq \bar{0}$ |
| (Q2) | $S(x) = S(y) \rightarrow x = y$ |
| (Q3) | $x \neq \bar{0} \rightarrow (\exists y)(x = S(y))$ |
| (Q4) | $x + \bar{0} = x$ |
| (Q5) | $x + S(y) = S(x + y)$ |
| (Q6) | $x * \bar{0} = \bar{0}$ |
| (Q7) | $x * S(y) = (x * y) + x$ |
| (Q8) | $x \leq y \equiv (\exists z)(z + x = y)$ |

Q is often called *Robinson arithmetic*. Note that thanks to our notational conventions, (Q7) may be written equally well as $x * S(y) = xy + x$ (omitting the parentheses and $*$ on the right hand side); but since we are beginning to develop axiomatic systems of arithmetic, we shall be slightly pedantic for some time. Later we shall again freely use our conventions. *Peano arithmetic* results from Q by adding the induction schema

$$\varphi(\bar{0}) \ \& \ (\forall x)(\varphi(x) \rightarrow \varphi(S(x))) \rightarrow (\forall x)\varphi(x).$$

This is indeed a *schema*: for each formula φ we have an induction axiom. Note that φ may contain free variables distinct from x as parameters. Peano arithmetic is denoted PA .

1.2 Lemma. In PA , the axiom (Q3) is redundant.

Proof. Let $\varphi(x)$ be $x = \bar{0} \vee (\exists y)(x = S(y))$ and proceed in PA : $\varphi(\bar{0})$ is obvious and $\varphi(S(x))$ too; thus we have $(\forall x)(\varphi(x) \rightarrow \varphi(S(x)))$, and thus $(\forall x)\varphi(x)$. $\square$

1.3. Particularly important fragments of PA result by restricting the induction schema to formulas φ from a prescribed class. This will be investigated in details in Sect. 2; here we make only a few particular choices. $I_{\text{open}}, I\Sigma_0, I\Sigma_1$ will denote the theory Q plus the induction schema for φ open, Σ_0, Σ_1 respectively. (We shall also investigate a theory with an extended language.) Note that in Part A we shall develop mainly theories containing $I\Sigma_1$ (and contained in PA). This is because in $I\Sigma_1$ we can formalize a proof of the fact that total Δ_1 functions are closed under primitive recursion (a careful formulation is presented below). This is the most important feature of fragments containing $I\Sigma_1$ and makes them remarkably different from weaker systems. Note also at this time that Chap. V deals with $I\Sigma_0$ and related theories and elaborates their specific problems. I_{open} will play only a marginal role in this book.

1.4. Note that by (Q3), each non-zero number x has a predecessor, i.e. a y such that $S(y) = x$. Thus we may define, in Q , a total function P by the following definition:

$$y = P(x) \equiv .(x = \bar{0} \& y = \bar{0}) \vee (x \neq \bar{0} \& S(y) = x).$$

We shall now prove several formulas in Q . Recall that for $m \in N$, $\bar{m}$ is the m -th numeral (cf. 0.28).

1.5 Lemma. The following formulas are provable in Q :

- (1) $x + y = \bar{0} \rightarrow . x = \bar{0} \& y = \bar{0},$
- (2) $x * y = \bar{0} \rightarrow . x = \bar{0} \vee y = \bar{0},$
- (3) $x + \bar{1} = S(x),$
- (4) $\bar{0} \leq x,$
- (5) $S(x) \leq \overline{n+1} \rightarrow x \leq \bar{n},$
- (6) $S(x) + \bar{n} = x + \overline{n+1},$
- (7) $\bar{n} \leq x \rightarrow . x = \bar{n} \vee \overline{n+1} \leq x.$

Proof. Proceed in Q . We prove (1)–(4). Take (1). If $y \neq \bar{0}$ then $y = S(z)$ for some z , thus $x + y = S(x + z) \neq \bar{0}$. If $x \neq \bar{0} \& y = \bar{0}$ then $x + y = S(z)$ for some z . This proves (1). Ad (2): assume $x, y \neq \bar{0}, x = S(u), y = S(v)$. Then $x * y = S(u) * S(v) = (S(u) * v) + S(u) = S(S(u) + v) \neq \bar{0}$. (3) is trivial.

(4) is obvious by (Q4). (5): If $z + S(x) = \overline{n+1}$ then $S(z + x) = S(\bar{n})$, thus $z + x = \bar{n}$. Note that (5) is a schema; for each n we have a proof. Also (6) is a schema; we shall construct the desired proofs by induction. Observe that we shall use *no* induction *within* the proofs (since we have no induction in Q); we shall construct the $(n+1)$ -th proof from the n -th one. This will often be the case. For $n = 0$, Q proves $S(x) + \bar{0} = S(x) = x + \bar{1}$. Assuming

(6) we get $Q \vdash S(x) + \overline{n+1} = S(x) + S(\bar{n}) = S(S(x) + \bar{n}) = S(x + \overline{n+1}) = x + S(\overline{n+1}) = x + \overline{n+2}$.

(7) In Q , assume $\bar{n} \leq x$ & $x \neq \bar{n}$; then, for some $z \neq \bar{0}$, $z + \bar{n} = x$. By (6), we get $x = P(z) + \overline{n+1}$, thus $\overline{n+1} \leq x$. $\square$

1.6 Theorem. For each $n, m \in N$, Q proves the following:

- (1) $\overline{m} + \bar{n} = \overline{m+n}$,
- (2) $\overline{m} * \bar{n} = \overline{m * n}$,
- (3) $\overline{m} \neq \bar{n}$ for $m \neq n$,
- (4) $x \leq \bar{n} \equiv . x = \bar{0} \vee x = \bar{1} \vee \dots \vee x = \bar{n}$,
- (5) $x \leq \bar{n} \vee \bar{n} \leq x$.

Proof. (1) We prove $Q \vdash \overline{m} + \bar{n} = \overline{m+n}$ by induction on n . For $n = 0$ we have to prove $Q \vdash \overline{m} + \bar{0} = \overline{m}$, which follows by (Q4). Assume we already have a proof of (1) and proceed in Q : $\overline{m} + \overline{n+1} = \overline{m} + S(\bar{n}) = S(\overline{m} + \bar{n}) = \overline{m+n+1}$. The proof of (2) is similar.

(3) Next we show that $m \neq n$ implies $Q \vdash \overline{m} \neq \bar{n}$. It suffices to assume $n < m$. For $m = 0$ the assumption is vacuous. Assume the assertion for m and let $n < m+1$. Then either $n = 0$ and (Q1) gives $Q \vdash \bar{n} \neq \overline{m+1}$ or $n = n_0 + 1$ and we have $Q \vdash \bar{n}_0 \neq \overline{m}$ by the inductive assumption; hence $Q \vdash \bar{n} \neq \overline{m+1}$ by (Q2).

(4) We construct proofs of the formulas in question by induction on n . For $n = 0$ see 1.5(1). Assume the assertion for n and consider $n+1$. The implication $\leftarrow$ is clearly provable using (1); thus proceed in Q and assume $x \leq \overline{n+1}$. If $x = \bar{0}$ we are done; therefore assume $x \neq \bar{0}$. By 1.5(5) we get $P(x) \leq \bar{n}$, thus $P(x) = \bar{0} \vee \dots \vee P(x) = \bar{n}$, which implies $x = \bar{1} \vee \dots \vee x = \overline{n+1}$.

(5) $Q \vdash \bar{0} \leq x$ by 1.5(4). Assume $Q \vdash \bar{n} \leq x \vee x \leq \bar{n}$ and proceed in Q . If $x \leq \bar{n}$ then $x \leq \overline{n+1}$ using (4) and (1); if $\bar{n} \leq x$ then, by 1.5(7), either $\bar{n} = x$, thus $x \leq \overline{n+1}$, or $\overline{n+1} \leq x$. $\square$

1.7 Remark. (1) $Q \vdash x + \bar{n} = \bar{k} \rightarrow x = \overline{k-n}$ (for $n \leq k$); this follows by iterated use of (Q2).

(2) Q proves

$$\vdash x + y = \bar{k} \rightarrow \bigvee_{i+j=k} x = \bar{i} \& y = \bar{j}$$

(by 1.6(4) using (1)).

1.8 Theorem. (Σ_1 -completeness of Q .) Let $\varphi(x)$ be a Σ_0 -formula with the only free variable x and let $N \models (\exists x)\varphi(x)$. Then $Q \vdash (\exists x)\varphi(x)$.

Proof. It is sufficient to show for each $\varphi(x_1, \dots, x_n) \in \Sigma_0$ that $N \models \varphi(\bar{k}_1, \dots, \bar{k}_n)$ implies $Q \vdash \varphi(\bar{k}_1, \dots, \bar{k}_n)$. First show, using 1.6(1),(2), that

for each term $t(x_1, \dots, x_n)$ and each n -tuple $k_1, \dots, k_n$ of elements of N ,

$$Q \vdash t(\overline{k_1}, \dots, \overline{k_n}) = \overline{Val(t(\overline{k_1}, \dots, \overline{k_n}))}$$

(thus, e.g. $Q \vdash (\overline{3} + \overline{5}) * \overline{8} = \overline{64}$). From this it follows, again using 1.6, that our assertion holds for φ atomic and negated atomic. (Observe that if $N \models \neg(\overline{k} \leq \overline{m})$ then $m < k$ and, by 1.6, $Q \vdash \overline{k} \leq \overline{m} \rightarrow (\overline{k} = \overline{0} \vee \dots \vee \overline{k} = \overline{m})$, thus $Q \vdash \neg(\overline{k} \leq \overline{m})$.) The induction step for logical connectives is easy. Finally, assume φ to be $(\exists y \leq x_1) \psi(y, x_1, \dots, x_n)$ and $N \models \varphi(\overline{k_1}, \dots, \overline{k_n})$; thus for some $k_0 \leq k_1$, $N \models \psi(\overline{k_0}, \overline{k_1}, \dots, \overline{k_n})$ and, by the induction hypothesis, $Q \vdash \psi(\overline{k_0}, \dots, \overline{k_n})$. This gives $Q \vdash \varphi(\overline{k_1}, \dots, \overline{k_n})$. Similarly for $\neg\varphi$, i.e. for $(\forall y \leq x) \neg\psi(y, x_1, \dots, x_n)$. $\square$

1.9 Remark. Thus each theory containing Q is Σ_1 -complete. (We shall show in Part B that no axiomatized consistent theory containing Q is Π_1 -complete.)

1.10 Theorem. The following formulas are provable in I_{open} :

- (1) $x + y = y + x$
- (2) $x + (y + z) = (x + y) + z$
- (3) $x * y = y * x$
- (4) $x * (y + z) = x * y + x * z$
- (5) $x * (y * z) = (x * y) * z$
- (6) $x + y = x + z \rightarrow x = y$
- (7) $x \leq y \vee y \leq x$
- (8) $x \leq y \& y \leq x \rightarrow x = y$
- (9) $(x \leq y \& y \leq z) \rightarrow x \leq z$
- (10) $x \leq y \equiv x + z \leq y + z$
- (11) $z \neq \overline{0} \& x * z = y * z \rightarrow x = y$
- (12) $z \neq \overline{0} \rightarrow (x \leq y \equiv x * z \leq y * z)$

Proof. We shall now use the induction schema inside I_{open} . At the beginning we shall give detailed proofs; later we shall omit details. The important thing is always to be sure that we use an instance of the induction schema given by a formula belonging to the class for which it is assumed; in our case, an open formula of the language L_0 .

(1) We first prove $(\forall x)(\overline{0} + x = x)$ in I_{open} . We use the induction axiom given by the open formula $\overline{0} + x = x$; denote it by $\varphi(x)$. First, $\varphi(\overline{0})$, i.e. $\overline{0} = \overline{0}$ follows by (Q3). To prove $(\forall x)(\varphi(x) \rightarrow \varphi(S(x)))$, assume $\overline{0} + x = x$ and compute as follows: $\overline{0} + S(x) = S(\overline{0} + x) = S(x)$. Thus by the induction axiom we get $(\forall x)\varphi(x)$.

Second, we prove $(\forall y)(S(x) + y = S(x + y))$. Let $\varphi(y)$ be $S(x) + y = S(x + y)$. The proof of $\varphi(\bar{0})$ is easy. Assume $\varphi(y)$ and prove $\varphi(S(y))$ as follows: $S(x) + S(y) = S(S(x) + y) = S(S(x + y)) = S(x + S(y))$. Thus we get $(\forall y)(S(x) + y = S(x + y))$. Compare this proof with the proof of 1.5(6): There we constructed, by metamathematical induction, infinitely many proofs (for each n , we constructed a proof of $S(x) + \bar{n} = x + \overline{n+1}$ in Q); here we have a single proof in I_{open} of $(\forall y)(S(x) + y = S(x + y))$. Clearly the latter formula implies each instance of the former schema. Now let us prove, in I_{open} , $(\forall x)(x + y = y + x)$. Let $\varphi(x)$ be $x + y = y + x$; we shall apply induction for φ . We have proved $\bar{0} + y = y + \bar{0}$; assume $x + y = y + x$ and reason as follows:

$$S(x) + y = S(x + y) = S(y + x) = y + S(x).$$

Thus we have proved $(\forall x)(\varphi(x) \rightarrow \varphi(S(x)))$; by the induction axiom we get $(\forall x)\varphi(x)$.

(2) We prove $(x + y) + z = x + (y + z)$ by induction on z . First, $(x + y) + \bar{0} = x + (y + \bar{0}) = x + y$ is clear. Assume $(x + y) + z = x + (y + z)$ and consider $(x + y) + S(z)$. We get $(x + y) + S(z) = S((x + y) + z) = S(x + (y + z)) = x + S(y + z) = x + (y + S(z))$. This completes the proof of (2). Note that from now on we may write sums like $x + y + z + u$ without parentheses.

(3) First prove $\bar{0} * x = \bar{0}$ by induction on x ; then prove $S(x) * y = (x * y) + y$ by induction on y ; finally, prove $x * y = y * x$ by induction on x . (Let us elaborate on the induction step for the second proof; assume $S(x) * y = (x * y) + y$. Then

$$\begin{aligned} S(x) * S(y) &= S(x) * y + S(x) && \text{(axiom (Q7))} \\ &= (x * y) + y + S(x) && \text{(inductive assumption} \\ &&& \text{plus associativity)} \\ &= (x * y) + S(y + x) && \text{(axiom (Q5))} \\ &= (x * y) + S(x + y) && \text{(commutativity (1))} \\ &= x * y + x + S(y) && \text{(axiom (Q5))} \\ &= x * S(y) + S(y) && \text{(axiom (Q7)} \\ &&& \text{plus associativity).)} \end{aligned}$$

(4) Prove $(x + y) * z = (x * z) + (y * z)$ by induction on z .

(5) Prove $(x * y) * z = x * (y * z)$ by induction on z . Thus products like $x * y * z * u$ (or $xyz u$) are meaningful.

(6) Prove $x + z = y + z \rightarrow x = y$ by induction on z . The induction step: assume $x + z = y + z \rightarrow x = y$ and $x + S(z) = y + S(z)$. Then $S(x + z) = S(y + z)$ by (Q5) and $x + z = y + z$ by (Q2); thus $x = y$.

(7) Prove $x \leq y \vee y \leq x$ by induction on x . (See the proof of 1.6(5).)

(8) Assume $x \leq y$ & $y \leq x$; thus $y = x + u$ and $x = y + v$. Then $x = x + u + v$ and $x = x + \bar{0}$; by (6), $u + v = \bar{0}$ and by 1.5(1), $u = v = \bar{0}$. Thus $x = y$.

(9) Easy using commutativity and associativity of addition.

(10) Use the same plus cancellation.

(11) We have to prove $z \neq \bar{0} \& x \neq y \rightarrow x * z \neq y * z$. By (7), we may assume $x \leq y$. Thus we prove the following by induction on z :

$$x \leq y \& x \neq y \& z \neq \bar{0} \rightarrow x * z \leq y * z \& x * z \neq y * z.$$

Observe that this is an open formula of L_0 ; we may write it as follows:

$$(*) \quad x < y \& z \neq \bar{0} \rightarrow x * z < y * z$$

(see the next remark). Now $x < y$ means $y = x + u$ for some $u \neq \bar{0}$. Nothing need be proved for $z = \bar{0}$. Assume $(*)$ and consider $S(z)$. We may assume $z \neq \bar{0}$, hence $x * z < y * z$, which means $y * z = x * z + v$ for some $v \neq \bar{0}$. Thus

$$y * S(z) = y * z + y = x * z + x + u + v = x * S(z) + (u + v)$$

and clearly $u + v \neq \bar{0}$; we get $x * S(z) < y * S(z)$.

(12) The proof of $x \leq y \rightarrow x * z \leq y * z$ is easy. Recall $(*)$ in (11) above and observe that, by using (8), $\neg(x \leq y)$ is equivalent to $y < x$. Thus we get

$$z \neq \bar{0} \rightarrow \neg(x \leq y) \rightarrow \neg(x * z \leq y * z),$$

i.e.

$$z \neq \bar{0} \rightarrow (x \leq y \equiv x * z \leq y * z).$$

□

1.11 Remark. (1) We have seen that I_{open} proves that $\leq$ is a linear order with $\bar{0}$ as the least element and that addition is monotone as well as multiplication by a non-zero element. As we previously did above, we define $x < y$ as $x < y \& x \neq y$ (the corresponding strict ordering); $x < y$ is equivalent to $(\exists u)(u \neq \bar{0} \& x + u = y)$. Furthermore, the ordering is discrete since for each x , $S(x)$ is the upper neighbour of x (and each non-zero element has a predecessor).

(2) I_{open} also proves $x * \bar{1} = x$ and, for each n , it proves

$$x * \bar{n} = (\dots((x + x) + \dots) + x \\ n \text{ times.}$$

The easy proofs are left to the reader as an exercise.

1.13 Definition (I_{open}). (1) x divides y (notation: $x \mid y$) if $(\exists z)(xz = y)$. (Recall that this is shorthand for $(\exists z)(x * z = y)$.)

(2) x is *prime* ($\text{Prime}(x)$) if $x > \bar{1}$ and $(\forall u)(u \mid x \rightarrow u = \bar{1} \vee u = x)$.

(3) $z = x - y$ if $(x \geq y \text{ and } x + z = y)$ or $(x < y \text{ and } z = \bar{0})$.

1.14 Lemma (I_{open}). (1) $x \mid y$ iff $(\exists z \leq y)(xz = y)$

(2) $\text{Prime}(x)$ iff $x > \bar{1}$ & $(\forall u \leq x)(u \mid x \rightarrow .u = \bar{1} \vee u = x)$.

Proof. (1) Assume $x \mid y$. If $y = \bar{0}$ then for $z = \bar{0}$ we have $z \leq y$ & $xz = y$; so assume $y \neq \bar{0}$ and $xz = y$. Then $x \neq \bar{0}$, thus $x \geq \bar{1}$ and $z = z * \bar{1} \leq z * x = y$.

(2) Observe that for $x \geq \bar{1}$, $u \mid x$ implies $u \leq x$. $\square$

1.15 Lemma (I_{open}).

(1) $\bar{1} \mid x$, $x \mid x$, $x \mid \bar{0}$, $(x \mid y \text{ and } y \mid z) \rightarrow x \mid z$;

(2) $(x \mid y \text{ \& } y \mid x) \rightarrow x = y$, $x \mid y \rightarrow x \mid yz$;

(3) $(x \mid y \text{ \& } x \mid z) \rightarrow x \mid (y + z)$;

(4) $y \neq \bar{0} \rightarrow (\exists! u \leq x)(\exists! v < y)(x = yu + v)$ (division with a remainder).

Proof. (1)–(3) are easy. We prove (4). Let $\varphi(u)$ be the formula $yu \leq x$. We have $\varphi(\bar{0})$ and for some u , e.g. for $u = x + 1$, we have $\neg\varphi(u)$ (since $y(x + 1) = yx + y \geq x + y > x$). By I_{open} , there is a u such that $yu \leq x$ and $y(u + 1) > x$. Since $\leq$ is a linear order and multiplication by a non-zero number is monotone, u is unique; furthermore $u \leq x$. Put $v = x - yu$; then $x = yu + v$ and $v < y$ (otherwise $v = y + v'$, $yu + y = y(u + 1) \leq x$, a contradiction.) Clearly, v is uniquely determined. $\square$

1.16 Lemma (I_{open}).

(1) If $x \geq y$ then $x - y = (x + z) - (y + z)$;

(2) if $x \geq y$ then $(x - y) + z = (x + z) - y$;

(3) $(x - y)z = xz - yz$;

(4) $z \mid x$ and $z \mid y$ implies $z \mid x - y$.

Proof. Exercise. $\square$

1.17 Definition and Lemma (I_{open}). x is *even* if $2 \mid x$; x is *odd* if it is not even. For each x , either x or $x + 1$ is even.

Proof. 0 is even; assume $x \neq 0$. Consider the open formula $2u \leq x$ and denote it by $\varphi(u)$. Clearly, $\varphi(0)$ and for some u , $\neg\varphi(u)$ (e.g. for $u = x$). Thus there is an u such that $2u \leq x$ and $2(u + 1) > x$. Then either $2u = x$ or $2u = x - 1$, and in the latter case $2(u + 1) = x + 1$. $\square$

1.18 Theorem and Definition ($\text{pairing}, I_{\text{open}}$). For each x, y , there is a unique z such that $2z = (x + y)(x + y + 1) + 2x$; this z is denoted (x, y) . For each z there is a unique pair x, y such that $z = (x, y)$.

Proof. Either $2 \mid (x + y)$ or $2 \mid (x + y + 1)$; thus $2 \mid ((x + y)(x + y + 1) + 2x)$. Therefore for some z , $2z = (x + y)(x + y + 1) + 2x$; clearly, this z is unique. Now take any z and using I_{open} , find an r such that $r(r + 1) \leq 2z$ and

$(r+1)(r+2) > 2z$. (Consider $\varphi(r) \equiv r(r+1) \leq 2z$.) Clearly, $2z - r(r+1)$ is even; let x be such that $2x = 2z - r(r+1)$. We have $x \leq r$ (otherwise we would have $2x > 2r$, $2r < 2z - r(r+1)$, $r^2 + 3r < 2z$ and $r^2 + 3r$ is even; thus $(r+1)(r+2) = r^2 + 3r + 2 \leq 2z$, a contradiction). Put $y = r - x$; then $x + y = r$, $2z = (x+y)(x+y+1) + 2x$, thus $z = (x, y)$. $\square$

1.19 Definition. Let T be a theory containing I_{open} . A formula $\varphi(\mathbf{x})$ is said to be Σ_n (Π_n) in T if there is a Σ_n formula (Π_n formula) $\psi(\mathbf{x})$ such that $T \vdash \varphi(\mathbf{x}) \equiv \psi(\mathbf{x})$. Furthermore, $\varphi(\mathbf{x})$ is said to be Δ_n in T if it is both Σ_n in T and Π_n in T .

1.20 Examples. $x \mid y$, $\text{Even}(x)$, $\text{Prime}(x)$, $z = (x, y)$ are Σ_0 in I_{open} .

1.21 Theorem. Let $T \supseteq I_{\text{open}}$. For each natural number n and any formulas φ, ψ :

- (1) if φ, ψ are $\Sigma_n(\Pi_n, \Delta_n)$ in T then so are $\varphi \& \psi$ and $\varphi \vee \psi$;
- (2) if φ is Δ_n in T then so is $\neg\varphi$;
- (3) if $n > 0$ and ψ is Σ_n in T then so is $(\exists x)\psi$;
- (4) if $n > 0$ and ψ is Π_n in T then so is $(\forall x)\psi$.

Proof. Fully analogous to that of 0.34. (For each n and each choice of φ, ψ , a T -proof is constructed.) $\square$

*

Now we turn to $I\Sigma_0$. Since $I\Sigma_0$ and related theories will be investigated in detail in Chap. V, we prove (or state) only some few basic facts.

1.22 Theorem. (1) $I\Sigma_0$ proves the *least number principle* for Σ_0 formulas: i.e. for each Σ_0 formula $I\Sigma_0$ proves

$$(\exists x)(\varphi(x) \rightarrow (\exists x)(\varphi(x) \& (\forall y < x) \neg \varphi(y))).$$

(2) For each Σ_0 formula, $I\Sigma_0$ proves the following *order induction*:

$$(\forall x)((\forall y < x)\varphi(y) \rightarrow \varphi(x)) \rightarrow (\forall x)\varphi(x).$$

Proof. (1) Assume $(\exists x)(\varphi(x) \& (\forall x)(\varphi(x) \rightarrow (\exists y < x)\varphi(y)))$ and apply induction to the formula $(\forall y < x) \neg \varphi(y)$ to obtain a contradiction. (2) Apply induction to the Σ_0 formula $(\forall y < x)\varphi(y)$. $\square$

1.23 Definition ($I\Sigma_0$). If $x, y \neq 0$ then $\text{gcd}(x, y)$ is the maximal u such that $u \mid x$ and $u \mid y$; otherwise $\text{gcd}(x, y) = 0$ (greatest common divisor).

Note that $\text{gcd}(x, y)$ exists since it is the least $u \leq \min(x, y)$ such that $(\forall v \leq \min(x, y))(u + 1 \leq v \rightarrow \neg(v \mid x \& v \mid y))$.

1.24 Lemma ($I\Sigma_0$). $0 < y \leq x \rightarrow (\exists u \leq x)(\exists v \leq x)(gcd(x, y) = xu - yv)$.

Proof. (0) Clearly we may assume $y < x$. Observe that if $0 < t < y$ and $t = xu - yv$ then there are $u < y$ and $v < x$ such that $t = xu - yv$. (To see this first show $u > y \equiv v > x$; thus if $u > y$ then we can replace u by $u - y$ and v by $v - x$. Apply $L\Sigma_0$.)

(1) Let z be the least number such that, for some $u, v < x$, $xu > yv$ & $z = xu - yv$. We show $z = gcd(x, y)$. First prove $z \mid x$. By 1.15(4), let $x = zs + t$, $t < z$. Clearly, $s \neq 0$ since $t < z < x$. If $t \neq 0$ then $t = x - zs = x - (xu - yv)s = x(1 + yq - u) - y(xq - v)$ where q is the least number such that $1 + yq > u$ and $xq > v$. Then clearly $1 + yq - u \leq x$, $xq - v \leq x$ and we get a contradiction with the minimality of z . Thus $t = 0$ and $z \mid x$.

(2) The proof of $z \mid y$ is similar (note that $z \leq y$ since $y = xy - y(x - 1)$).

(3) Now if $w \mid x$ and $w \mid y$ then $w \mid z$ (by 1.13(4)); thus $w \leq z$ and $z = gcd(x, y)$. This completes the proof. $\square$

1.25 Lemma ($I\Sigma_0$). $gcd(x, y) = 1 \rightarrow (\forall z)(x \mid yz \rightarrow x \mid z)$.

Proof. Assume $gcd(x, y) = 1$ and $y \leq x$. Then, for some $u, v \leq x$, $1 = xu - yv$. Now if $x \mid yz$ then $x \mid xuz - yvz$, $x \mid (xu - yv)z$, $x \mid z$. $\square$

1.26 Lemma ($I\Sigma_0$). (1) For each $x > 1$, there is a $y \leq x$ such that $Prime(y)$ and $y \mid x$. (2) If $Prime(x)$ and $x \mid yz$ then $x \mid y$ or $x \mid z$.

Proof. (1) follows by the least number principle (1.22). (2) Assume that x does not divide y ; since x is prime this means $gcd(x, y) = 1$. Thus $x \mid z$ by 1.25. $\square$

1.27 Remark. The reader may move now to Chap. V where $I\Sigma_0$ is investigated in details. He will find there, among other things, a proof of the following theorem claiming that in $I\Sigma_0$ exponentiation is Σ_0 definable as a possibly partial function:

There is a Σ_0 formula $exp(z, x, y)$ such that $I\Sigma_0$ proves the following:

- (1) $exp(z_1, x, y) \& exp(z_2, x, y) \rightarrow z_1 = z_2$,
- (2) $exp(1, x, 0)$,
- (3) $exp(z, x, y) \rightarrow exp(zx, x, S(y))$,
- (4) $exp(z, x, y) \& y' < y \rightarrow (\exists z') exp(z', x, y')$.

Many results of Chaps. I-IV are independent of this theorem; the reader may postpone reading its proof. In subsection (c) of the present section, we shall prove a weaker (and classical) result saying that there is a formula exp which is Δ_1 in $I\Sigma_1$ and such that $I\Sigma_1$ proves (1)-(4). This weaker result is basic for Chaps. I-IV.

(b) Coding Finite Sets and Sequences; the Theory $I\Sigma_0(exp)$

1.28. In this subsection we shall investigate a theory stronger than $I\Sigma_0$ and having a richer language: we extend the language by a new unary function symbol $\bar{2}^x$ for the x -th power of two. The extended language is denoted $L_0(exp)$. $\Sigma_0(exp)$ formulas result from atomic formulas of $L_0(exp)$ by iterated application of logical connectives and bounded quantifiers of the form $(\forall x \leq y)$, $(\exists x \leq y)$. The theory $I\Sigma_0(exp)$ has the following axioms:

- (1) axioms of Q ,
 (2) two axioms for exponentiation, namely:

$$\begin{aligned}\bar{2}^{\bar{0}} &= \bar{1}, \\ \bar{2}^{S(x)} &= \bar{2}^x * \bar{2},\end{aligned}$$

- (3) induction schema for all $\Sigma_0(exp)$ formulas.

1.29. We shall need another class of formulas called $\Sigma_0^{exp}(exp)$ formulas: they result from atomic $L_0(exp)$ formulas by iterated application of logical connectives and bounded quantifiers of the following form:

$$(\forall x \leq y), \quad (\exists x \leq y), \quad (\forall x \leq \bar{2}^y), \quad (\exists x \leq \bar{2}^y).$$

1.30 Theorem ($I\Sigma_0(exp)$). (1) $x < 2^x$; (2) $x < y \rightarrow 2^x < 2^y$; (3) $2^x * 2^y = 2^{x+y}$.

Proof. (1) is easy. (2) Use induction on y . Nothing is to be proved for $y = \bar{0}$. Assume $x < y \rightarrow 2^x < 2^y$ and $x < y + 1$.

Case 1: $x < y$ is false; then necessarily $x = y$, $2^x = 2^y > y \geq \bar{0}$, thus $2^{y+1} = 2^y * 2 = 2^y + 2^y > 2^y = 2^x$.

Case 2: $x < y$ is true; then $2^x < 2^y < 2^y * 2 = 2^{y+1}$. Thus the induction step is proved. (3) Induction on y . $\square$

1.30 Lemma. The schema of induction for $\Sigma_0^{exp}(exp)$ formulas is derivable in $I\Sigma_0(exp)$.

Proof. Clearly, it is enough to prove the least number principle for each $\Sigma_0^{exp}(exp)$ formula. We claim that for each such formula $\varphi(\mathbf{x})$ there is a $\Sigma_0(exp)$ formula $\varphi_0(\mathbf{x}, y)$ and a term $t(\mathbf{x})$ of $L_0(exp)$ such that $I\Sigma_0(exp)$ proves the following:

$$(*) \quad (\forall y \geq t(\mathbf{x}))(\varphi(\mathbf{x}) \equiv \varphi_0(\mathbf{x}, y))$$

This is evident for φ atomic ($t(\mathbf{x}) = \bar{0}$, $\varphi_0(\mathbf{x}, y) = \varphi(\mathbf{x})$). If $(*)$ holds for φ, φ_0, t then it holds for $\neg\varphi, \neg\varphi_0, t$; if it holds for $\varphi_1, \varphi_{10}, t_1$ and $\varphi_2, \varphi_{20}, t_2$ then it

holds for φ_1 & φ_2, φ_{10} & $\varphi_{20}, t_1 + t_2$. Assume that we have $\varphi(\mathbf{x}, u), \varphi_0(\mathbf{x}, u, y)$ and $t(\mathbf{x}, u)$ satisfying the analogue of (*) and investigate $\psi(\mathbf{x})$ being $(\forall u \leq 2^x)\varphi(\mathbf{x}, u)$ (where $\mathbf{x} = x, \dots$). Then (*) holds for $\psi'_0(\mathbf{x})$ being $(\forall u \leq y)(u \leq 2^x \rightarrow \varphi_0(\mathbf{x}, u, y))$ and $t'(\mathbf{x})$ being $2^x + t(\mathbf{x}, 2^x)$. (Note that $I\Sigma_0(exp)$ proves $u \leq 2^x \rightarrow t(\mathbf{x}, u) \leq t(\mathbf{x}, 2^x)$.) This proves our claim.

To complete the proof of 1.30, let $\varphi, \varphi_0 \in \Sigma_0^{exp}(exp)$ be as in the claim above. In $I\Sigma_0(exp)$, assume $\varphi(\mathbf{x})$ and put $y = t(\mathbf{x})$. Then we have $\varphi_0(\mathbf{x}, y)$; write it $\varphi_0(x_1, x_2, \dots, y)$. By the least number principle for $I\Sigma_0(exp)$ formulas, let x'_1 be the least x_1 such that $\varphi_0(x_1, x_2, \dots, y)$; this x'_1 is the least x_1 such that $\varphi(x_1, x_2, \dots)$. (Note that $t(x'_1, x_2, \dots) \leq t(x_1, x_2, \dots)$.)

1.31 Lemma and Definition ($I\Sigma_0(exp)$). For each x, y , there are unique $u \leq y, v \leq 1, w < 2^x$ such that

$$y = 2^{x+1} * u + 2^x * v + w.$$

The unique $v \leq 1$ such that $(\exists u \leq y)(\exists w < 2^x)(y = 2^{x+1} * u + 2^x * v + w)$ is called the x -th bit of y and denoted $bit(x, y)$. We further define

$$x \in y \equiv bit(x, y) = 1.$$

(Note that $x \in y$ is $\Sigma_0(exp)$ in $I\Sigma_0(exp)$.)

Proof. By 1.15(4) (division with remainder), there are $u \leq y$ and $q < 2^{x+1}$ such that $y = 2^{x+1} * u + q$; by the same theorem, there are $v \leq q$ and $w < 2^x$ such that $q = 2^x * v + w$. The numbers u, v, w are uniquely determined. Moreover, u must be less than 2 since otherwise we would have $q \geq 2^{x+1}$, a contradiction. $\square$

1.32 Lemma ($I\Sigma_0(exp)$). (1) $x \in y \rightarrow x < y$. (2) $y = \bar{0} \rightarrow (\forall x)(x \notin y)$. (3) $y \neq \bar{0} \rightarrow y$ has a least and a largest element. (4) $y < 2^u \equiv (\forall x)(x \in y \rightarrow x < u)$.

Proof. (1) $bit(x, y) = 1$ implies $y \geq 2^x > x$. (2) follows from (1). Consider (3): If $y \neq \bar{0}$ then we first claim that there is a largest $x < y$ such that $2^x \leq y$; then it follows easily that $x \in y$ and x is the largest element of y . Having $(\exists x < y)(x \in y)$ we get a least element of y by the least number principle. Thus let us prove the claim. Let x be the least number such that $(\forall z < y)(2^z \leq y \rightarrow z \leq x)$. Then clearly $2^x \leq y$ and x is the largest such number. (4) The implication $\rightarrow$ is obvious. To prove the converse, assume $y \geq 2^u$ and let x be the largest such that $y \geq 2^x$. Then $y = 2^x + w$ for some $w < 2^x$, which implies $x \in y$; and obviously, $x \geq u$. $\square$

1.33 Definition ($I\Sigma_0(exp)$). $x \subseteq y \equiv (\forall u < x)(u \in x \rightarrow u \in y)$
(Note that by 1.32, $x \subseteq y$ is equivalent to $(\forall u)(u \in x \rightarrow u \in y)$.)

1.34 Lemma ($I\Sigma_0(exp)$). (1) If $w < 2^z$ then for all x ,

$$x \in 2^z + w \equiv .x \in w \vee x = z.$$

(2) For each u, y ,

$$y < 2^u \equiv y \subseteq 2^u - 1.$$

Proof. ($\leftarrow$) Clearly $z \in 2^z + w$ by definition. Assume $x \in w$, i.e. $w = 2^{x+1} * s + 2^x + t$, $t < 2^x$. Then $2^z + w = 2^z + 2^{x+1} * s + 2^x + t = 2^{x+1}(2^{z-x-1} + s) + 2^x + t$, i.e. $x \in 2^z + w$.

($\rightarrow$) Assume $x \in 2^z + w$; then $x \leq z$ (since $2^z + w \leq 2^{z+1}$). If $x = z$ then we are done; assume $x < z$. We have $2^z + w = 2^{x+1} * s + 2^x + t$.

Claim. $2^{x+1} * s \geq 2^z$. Otherwise $2^{x+1} * s < 2^z$ and since $2^z = 2^{x+1} * 2^{z-x-1}$, i.e. $2^{x+1} \mid 2^z$, we get $2^{x+1} * s \leq 2^z - 2^{x+1}$. But then $2^{x+1} * s + 2^x + t < 2^z$, a contradiction. The claim is proved.

Thus $s \geq 2^{z-x-1}$ and $s < 2^{z-x}$; this implies that in dividing s by 2^{z-x-1} we get $s = 2^{z-x-1} + q$ ($q < 2^{z-x-1}$),

$$\begin{aligned} 2^z + w &= 2^{x+1}(2^{z-x-1} + q) + 2^x + t = 2^z + 2^{x+1} * q + 2^x + t, \\ w &= 2^{x+1} * q + 2^x + t, \end{aligned}$$

and consequently $x \in w$. This completes the proof of (1).

(2) By 1.31, it suffices to show

- (i) $x < u \rightarrow \text{bit}(x, 2^u - 1) = 1$,
- (ii) $x \geq u \rightarrow \text{bit}(x, 2^u - 1) = 0$.

But (ii) is obvious; let us prove

$$(*) \quad (\forall u)(\forall x < u)(\text{bit}(x, 2^u - 1) = 1).$$

(Show that the formula following $(\forall u)$ is $\Sigma^{exp}(exp)$ in $I\Sigma_0(exp)$.) Nothing need be proved for $u = 0$. Assume $(*)$ and consider $u + 1$: $2^{u+1} - 1 = 2^u + (2^u - 1)$. Thus $\text{bit}(u, 2^{u+1} - 1) = 1$ and, by $(*)$, for $x < u$, $\text{bit}(x, 2^u - 1) = 1$. But by (1), $\text{bit}(x, 2^u - 1) = \text{bit}(x, 2^{u+1} - 1)$. $\square$

1.35 Definition ($I\Sigma_0(exp)$). For each x, y , put

$$\begin{aligned} \max(x) &= \text{the largest element of } x \text{ if } x \neq \bar{0}, \\ &= \bar{0} \text{ for } x = \bar{0} \text{ (cf. 1.34(3))} \\ (\leq x) &= 2^{x+1} - 1 \\ (< x) &= 2^x - 1 \\ \text{seg}(x) &= (\leq \max(x)). \end{aligned}$$

Note that, by 1.34(2), $u \in (\leq x)$ iff $u \leq x$. We further make the following definition: y is a *restriction* of x to z (in symbols: $Restrict(y, x, z)$) if $y < 2^z$ and $(\forall u < z)(u \in x \equiv u \in y)$. Note that this notion is $\Sigma_0^{exp}(exp)$ in $IS_0(exp)$ and that the latter theory proves the following: if $Restrict(y, x, z)$ and $z \geq 2^x$ then $(\forall u)(u \in x \equiv u \in y)$.

1.36 Theorem ($IS_0(exp)$ -comprehension). For each $\Sigma_0^{exp}(exp)$ -formula $\varphi(u, p)$, $IS_0(exp)$ proves

$$(\forall x)(\exists y < 2^x)(\forall u < x)(u \in y \equiv \varphi(u, p)).$$

Proof. We apply induction on x to the formula

$$(*) \quad (\exists y < 2^x)(\forall u < x)(u \in y \equiv \varphi(u, p)).$$

The case $x = 0$ is trivial. Assume $(*)$ and let y be as in $(*)$. Consider $x + 1$. If $\varphi(x, p)$ holds, then put $y' = y + 2^x$; by the preceding,

$$(\forall u < x + 1)(u \in y' \equiv (u < x \& \varphi(u, p) \vee u = x) \equiv \varphi(u, p)).$$

If $\neg\varphi(x, p)$ holds, then put $y' = y$ and argue similarly. □

1.37 Theorem ($IS_0(exp)$). If $x \subseteq y$ then $x \leq y$.

Proof. Assume $x \subseteq y$ and let us prove the following by induction on z :

$$(*) \quad (\forall x' < 2^z)(\forall y' < 2^z)(Restrict(x', x, z) \& Restrict(y', y, z) \rightarrow x' \leq y').$$

The case $z = 0$ is trivial ($x' = y' = 0$). Assume $(*)$ and consider $z + 1$. Let $Restrict(x'', x, z + 1), Restrict(y'', y, z + 1)$. Put

$$\begin{aligned} x' &= x'' - 2^z \text{ if } z \in x, \\ x' &= x'' \text{ if } z \notin x \end{aligned}$$

and similarly for y' . Then $Restrict(x', x, z), Restrict(y', y, z)$ and $x' \leq y'$. If $z \notin x$ then $x'' = x' \leq y' \leq y''$ by $(*)$. If $z \in x$ then $z \in y$ and $x'' = x' + 2^z \leq y' + 2^z = y''$ again by $(*)$. This completes the proof. □

1.38 Corollary ($IS_0(exp)$ -extensionality). If $(\forall u)(u \in x \equiv u \in y)$ then $x = y$. (In particular, the restriction of x to z is uniquely determined.)

1.39 Theorem ($I\Sigma_0(exp)$). For any x, y , there exist uniquely determined numbers $x \cup y, x \cap y, x - y, x \times y, \bigcup x, P(x)$ having the following properties:

- (union) $(\forall u)(u \in x \cup y \equiv . u \in x \vee u \in y)$
- (intersection) $(\forall u)(u \in x \cap y \equiv . u \in x \& u \in y)$
- (set difference) $(\forall u)(u \in x \setminus y \equiv . u \in x \& x \not\subseteq y)$
- (Cartesian product) $(\forall u)(u \in x \times y \equiv . (\exists u \in x)(\exists v \in y)(u = (v, w)))$
- (sum set) $(\forall u)(u \in \bigcup x \equiv . (\exists v \in x)(u \in v))$
- (power set) $(\forall u)(u \in P(x) \equiv . u \subseteq x)$

Proof. We always find a z such that the set in question is a $\Sigma_0^{exp}(exp)$ -definable subset of $seg(z)$. Take the union and let $x \leq y$. Then $seg(x) \subseteq seg(y)$ and we put $z = y$; there is a w such that $(\forall u)(u \in w \equiv u \in seg(z) \& (u \in x \vee u \in y))$ (by comprehension). Thus u is $x \cup y$. Similarly for $x \cap y, x - y, \bigcup x$. For $x \times y$ observe that $u < x \& v < y$ implies $(u, v) < (x + y + 1) * (x + y + 1)$ (say); thus there is a w such that

$$(\forall z)(z \in w \equiv . z \in seg((x + y + 1)^2) \& (\exists u \in x)(\exists v \in y)(z = (u, v))).$$

Finally for the power set, observe that $u \subseteq x$ implies $u \leq x$; thus there is a w such that

$$(\forall u)(u \in w \equiv . u \in seg(x + 1) \& (\forall v \in u)(v \in x)). \quad \square$$

Remark. Next we introduce some well-known notions concerning (finite) sets and show that their obvious properties are provable in $I\Sigma_0(exp)$. Note that in $I\Sigma_0(exp)$ each number is a finite set (codes a finite set); thus instead of saying "a finite set" we just say "a number"; but we call the reader's attention to properties definable using the membership predicate (as defined above).

1.40 Definition ($I\Sigma_0(exp)$).

(1) x is a *relation* if x is a set of ordered pairs:

$$Rel(x) \equiv (\forall u \in x)(\exists v, w \in u)(u = (v, w))$$

$$(\text{domain}) \quad v \in dom(x) \equiv (\exists u \in x)(\exists w \leq u)(u = (v, w))$$

$$(\text{range}) \quad w \in rng(x) \equiv (\exists u \in x)(\exists v \leq u)(u = (v, w))$$

(show that $y = dom(x), y = rng(x)$ are $\Sigma_0^{exp}(exp)$ in $I\Sigma_0(exp)$.)

(2) y is a *linear order* on x if $Rel(y), dom(y) = rng(y) = x$ and y is reflexive, transitive and dichotomic on x , i.e.

$$\begin{aligned} &(\forall \dots)((u, v) \in y \& (v, w) \in y. \rightarrow (u, w) \in y), \\ &(\forall u, v \in x)((u, v) \in y \vee (v, u) \in y), \\ &(\forall u, v \in x)((u, v) \in y \& (v, u) \in y. \rightarrow u = v). \end{aligned}$$

(3) y is a mapping of x into z if $Rel(y)$, $dom(y) = x$, $rng(y) \subseteq z$ and

$$(\forall u \in x)(\forall v, w \in z)((u, v) \in y \& (u, w) \in y \rightarrow v = w).$$

A mapping is an *injection* (or: is one-one) if

$$(\forall u_1, u_2 \in x)(\forall v \in z)((u_1, v) \in y \& (u_2, v) \in y \rightarrow u_1 = u_2);$$

y is a *bijection* of x onto z if, additionally, $rng(y) = z$.

(4) For each x , the *natural ordering* of x is the unique linear order y on x such that, for each $u, v \in x$, $(u, v) \in y \equiv u \leq v$. (Show the existence of y using comprehension.)

1.41 Theorem ($I\Sigma_0(exp)$). (1) (Cardinality.) For each x , there is a unique $y = card(x)$ such that there is a bijection of x to $(< y)$.

(2) (Pigeon-hole principle for finite sets.) If $card(x) < card(y)$ then there is no injection of y into x .

(3) If y is a linear order on $x \neq 0$ then x has a largest and a least element with respect to y .

Proof. One only checks that the usual proofs formalize in $I\Sigma_0(exp)$.

(1) The desired bijection f , if it exists, satisfies $f \leq 2^{(x+y+1)^2}$; thus we may prove the following by $\Sigma_0^{exp}(exp)$ -induction on x :

$$(*) \quad (\exists! y \leq x)(\exists f \leq 2^{(2x+1)^2})(f \text{ is a bijection of } x \text{ to } (< y)).$$

(Check that this is $\Sigma_0^{exp}(exp)$.)

For $x = 0$ we get $y = 0$; assume $(*)$ for each $x < z$ and investigate z . Let $u = \max(z)$ and $x = z - 2^u$; then $x < z$, $v \in x \equiv .v \in z \& v \neq u$ and there is a y satisfying $(*)$. Let f be the corresponding mapping; we extend f to a bijection g of z onto $(< y + 1)$ by defining $g(v) = f(v)$ for $v \in x$, $g(u) = y$. The mapping (relation, set) exists thanks to $\Sigma_0^{exp}(exp)$ -comprehension.

To prove both the uniqueness of cardinality and the pigeon-hole principle, we show the following by $\Sigma_0^{exp}(exp)$ -induction:

$$\neg(\exists f \leq 2^{(2x+1)^2})(f \text{ is an injection of } (< x + 1) \text{ into } (< x)).$$

This is clear for $x = 0$. Assume $(*)$ and let f be an injection of $(< x + 2)$ into $(< x + 1)$. We may also assume that $f(x + 1) = x$ (if not we change f for the arguments $x + 1$ and $f^{-1}(x)$ using comprehension). But then the restriction of f to $(< x + 1)$ is an injection of $(< x + 1)$ into $(< x)$, a contradiction. This completes the proof of the induction step.

Consequently, if there is a bijection of $(< x)$ onto $(< y)$ then $x = y$. Thus each finite set has its unique cardinality; and we get the pigeon-hole

principle using composition of mappings. The fact that the composition of two mappings that are finite sets is a mapping (and a finite set) follows by comprehension.

(3) By induction on x ; note that the universal quantifier $(\forall y)$ may be bounded by $2^{(2 \max(x)+1)^2}$ so that $\Sigma_0^{exp}(exp)$ -induction applies. Assume the assertion for all $z < x$ and consider z ; let u be the maximal element of z with respect to the ordering $<$ and let $v \in x \equiv .v \in z \& v \neq u$. Then $x < z$ and if $y' = y \cap (x \times x)$ we see that y' is a linear order on x . By the induction assumption, x has a maximal element u' with respect to y' . Then either u or u' is maximal in z with respect to y . $\square$

1.42 Definition ($I\Sigma_0(exp)$). Now we finally come to our definition of finite sequences; they are naturally defined as particular mappings.

$$\begin{aligned} Seq(z) &\equiv (\exists x \leq z)(z \text{ is a mapping} \& dom(z) = (< x)); \\ Seq(z) \rightarrow .lh(z) = x &\text{ iff } dom(z) = (< x); \\ Seq(z) \& u < lh(z) &\rightarrow (z)_u = v \text{ iff } (u, v) \in z; \\ Seq(z) \& u \geq lh(z) &\rightarrow (z)_u = 0; \\ Seq(z) \rightarrow .w = z \frown \langle x \rangle &\text{ iff } (\forall u < w)(u \in w \equiv .u \in z \vee u = (lh(z), x)); \\ \text{for } \neg Seq(z), \quad z \frown \langle x \rangle &= 0; \\ \emptyset \text{ is the empty sequence; } lh(\emptyset) &= 0; \\ \langle x \rangle = \emptyset \frown \langle x \rangle; \langle x_1, \dots, x_n, x_{n+1} \rangle &= \langle x_1, \dots, x_n \rangle \frown \langle x_{n+1} \rangle \\ (n = 1, 2, \dots). \end{aligned}$$

Note that all notions defined are $\Sigma_0^{exp}(exp)$ in $I\Sigma_0(exp)$.

1.43 Theorem. $I\Sigma_0(exp)$ proves the basic properties of sequences (as formulated in 0.41): For $Seq(s), Seq(s')$,

- (1) $lh(s) \leq s$ and $(\forall u < lh(s))((s)_u < s)$;
- (2) $Seq(0) \& lh(0) = 0$;
- (3) $(\forall u < lh(s))((s)_u = (s \frown \langle z \rangle)_u \& (s \frown \langle z \rangle)_{lh(s)} = z \& \\ \& lh(s \frown \langle z \rangle) = lh(s) + 1$;
- (4) if $lh(s) \leq lh(s')$ and $(\forall u < lh(s))((s)_u \leq (s')_u)$ then $s \leq s'$
- (5) $(\forall u)(\exists v > u) \neg Seq(v)$.

(Consequently, if s, s' have the same length and the same corresponding elements then $s = s'$.)

Proof. (2) is trivial: 0 is the empty mapping. All monotonicities of (1) and (4) follow from the monotonicity of the ordered pair $(x, y \leq (x, y))$ and of

membership ($x \in y \rightarrow x < y$). Indeed, if $s \neq 0$, $Seq(s)$ and $dom(s) = (< x)$, then $lh(s) = x$ and for some y , $(x - 1, y) \in s$. Thus $x - 1 \leq (x - 1, y) < s$ and hence $x < s$. If $(u, y) \in s$ then $y \leq (u, y) < s$. This proves (1)

We prove (4). First observe that for each sequence s and each $i < lh(s)$ we have the shortening of s to i (denoted by $s \downarrow i$ for a moment) which results from restricting the domain of s to $(< i)$. Existence follows by comprehension; and $s \downarrow lh(s)$ is s . Assume $lh(s) \leq lh(s')$ and $(\forall i < lh(s))((s)_i \leq (s')_i)$; let us prove $(\forall i < lh(s))((s \downarrow i) \leq (s' \downarrow i))$ by induction on i . Note that $s \downarrow (i + 1) = (s \downarrow i) + 2^{(i, s)_i}$. We get $s \leq s' \downarrow lh(s)$; but $s' \downarrow lh(s) \leq s'$ is obvious since $s' \downarrow lh(s) \subseteq s'$.

Proofs of (3) and (5) are easy. □

Remark. Since $I\Sigma_0(exp)$ proves $(x, y) \leq (x + y + 1)^2$, it also proves the following: if $lh(s) = z$ and $(\forall i < z)((s)_i \leq x)$ then $s \leq 2^{(x+z+1)^2}$.

(c) Provably Recursive Functions; the Theory $I\Sigma_1$

We are going to investigate the theory $I\Sigma_1$ (Q plus induction for Σ_1 formulas). We have two main goals in this subsection: first, to show that exponentiation is definable in $I\Sigma_1$ in such a way that $I\Sigma_0(exp)$ becomes a subtheory of $I\Sigma_1$; and, second, to show that the class of all Σ_1 definable functions (on N) whose totalness is provable in $I\Sigma_1$ is closed under primitive recursion and therefore each primitive recursive function is $I\Sigma_1$ -provably recursive. The first goal is reached by developing, in some extent, an alternative weak (non-extensional) coding of finite sets in $I\Sigma_1$ based on divisibility. This coding could be fully developed, but we use it only to define exponentiation; then it may be forgotten and the coding based on exponentiation may be used instead. To reach the second goal, we prove more than stated: we show that for each Σ_1 -definition of a total function in $I\Sigma_1$, we may construct another Σ_1 -definition of a total function in $I\Sigma_1$ such that $I\Sigma_1$ proves that the latter function is the primitive recursive iteration of the former one. This will be frequently used throughout the book (and will be strengthened in the next subsection).

We start with some basic observations on $I\Sigma_1$. Trivially, $I\Sigma_1$ contains $I\Sigma_0$ and therefore proves the least number principle for Σ_0 formulas. (We postpone the proof of the fact that $I\Sigma_1$ proves the least number principle for Σ_1 formulas to Sect. 2.) Now we show that $I\Sigma_1$ proves *collection* for Σ_0 formulas.

1.44 Theorem. For each Σ_0 formula $\varphi(x, y)$ (possibly containing parameters distinct from u, v),

$$I\Sigma_1 \vdash (\forall x < u)(\exists y)\varphi(x, y) \rightarrow (\exists v)(\forall x < u)(\exists y < v)\varphi(x, y).$$

Proof. We proceed in $I\Sigma_1$. Assume $(\forall x < u)(\exists y)\varphi(x, y)$ and prove, by Σ_1 induction on w , the following:

$$(*) \quad w \leq u \rightarrow (\exists v)(\forall x < w)(\exists y < v)\varphi(x, y).$$

Nothing is to be proved for $w = 0$; assuming $(*)$ for w and $w+1 \leq u$, let y_0 be such that $\varphi(w, y_0)$ and put $v' = \max(v, y')$. Then we get $(\forall x < w+1)(\exists y < v')\varphi(x, y)$. $\square$

1.45 Corollary. (1) $I\Sigma_1$ proves collection for Σ_1 formulas. (2) Formulas Σ_1 in $I\Sigma_1$ are closed under bounded quantification. (3) Thus formulas Δ_1 in $I\Sigma_1$ are closed under bounded quantification. (4) $I\Sigma_1$ proves the following *order induction* for each Σ_1 formula $\varphi(x)$ (cf. 1.22):

$$(\forall x)((\forall y < x)\varphi(y) \rightarrow \varphi(x)) \rightarrow (\forall x)\varphi(x).$$

Proof. (1) by contraction of quantifiers. (2) Let $\varphi(\mathbf{x})$ be $(\exists y)\varphi_0(\mathbf{x}, y)$ where φ_0 is Σ_0 and $\mathbf{x}$ is $x_1, x_2, \dots$. Then the formula $(\exists x_1 \leq x_2)\varphi(\mathbf{x})$ is Σ_1 in $I\Sigma_1$ (commute the existential quantifiers) and so is $(\forall x_1 \leq x_2)\varphi(\mathbf{x})$ since it is equivalent to $(\exists v)(\forall x_1 \leq x_2)(\exists y \leq v)\varphi_0(\mathbf{x}, y)$. (3) is immediate. (4) Apply induction to the formula $(\forall y < x)\varphi(y)$ observing that the last formula is Σ_1 in $I\Sigma_1$ by (2). $\square$

Now we shall exhibit an auxiliary coding of sequences.

1.46 Lemma and Definition ($I\Sigma_1$). For each x , there is a least y such that all positive $u \leq x$ divide y . We write $y = \text{hull}(x)$.

Proof. We prove $(\forall x)(\exists y)(\forall u \leq x)(0 < u \rightarrow u | y)$ by Σ_1 induction; the existence of a least such y follows by the least number principle for Σ_0 formulas.

For $x = 0$ the assertion is vacuous. Assume $(\forall u \leq x)(0 < u \rightarrow u | y)$ and take $y' = y * (x+1)$; then $(\forall u \leq x+1)(u > 0 \rightarrow u | y')$. $\square$

1.47 Definition. $x \in_0 (y, z)$ iff $(1 + (1+x) * z) | y$.

1.48 Lemma (Comprehension). For each Σ_0 formula $\varphi(u)$ $I\Sigma_1$ proves the following:

$$(\forall x)(\exists y, z)(\forall u < x)(u \in_0 (y, z) \equiv \varphi(u)).$$

(φ may contain free variables distinct from u, y, z as parameters.)

Proof. We may assume $x > 1$. Let $z = \text{hull}(x)$. We claim that for $u < v < x$ the numbers $1 + (1+u)z, 1 + (1+v)z$ are relatively prime, i.e. their greatest

common divisor is 1. (This claim is proved below.) Using this we show the following by Σ_1 induction on t :

$$(*) \quad (t \leq x \rightarrow (\exists y)(\forall u < x)[(u < t \& \varphi(u) \rightarrow u \in_0(y, z)) \& \\ \& (u \geq t \vee \neg \varphi(u))] \rightarrow \gcd(y, 1 + (1 + u)z) = 1].$$

For $t = 0$ take $y = 1$. Assume $(*)$ and consider $t + 1$. Let y be as in $(*)$.

Case 1: $\neg \varphi(t)$. Then $(*)$ holds for t replaced by $t + 1$ and for y as it stands.

Case 2: $\varphi(t)$. Put $y' = y * (1 + (1 + t)) * z$. Then clearly $t \in_0(y', z)$ and $u < t \& \varphi(u)$ implies $u \in_0(y', z)$. If $u > t \vee \neg \varphi(u)$ then $\gcd(y, 1 + (1 + u)z) = 1$ and $\gcd(y, 1 + (1 + t)z) = 1$. Moreover, by our claim, $\gcd(1 + (1 + u)z, 1 + (1 + t)z) = 1$. Now if c divides both $1 + (1 + u)z$ and $y' = y * (1 + (1 + t)) * z$ then, by 1.25, $c \mid y$ or $c \mid 1 + (1 + t)z$; thus $c = 1$ and we are done. It remains to prove the claim. $\square$

(*Proof of the claim.*) Assume $u < v < x$; then $0 < v - u < x$ and $v - u \mid z$. Write u_1, v_1 for $1 + u, 1 + v$ and let c be such that $c \mid 1 + u_1z, c \mid 1 + v_1z$. Then for some a, b we have $1 + u_1z = ac, 1 + v_1z = bc, (1 + u_1z) \mid abc = a(1 + v_1z)$. Since trivially $(1 + u_1z) \mid a(1 + u_1z)$ we get (by subtracting) $(1 + u_1z) \mid a(v_1z - u_1z) = a(v - u)z$. Evidently, $\gcd(1 + u_1z, z) = 1$; thus, by 1.25, $(1 + u_1z) \mid a(v - u) \mid az$. By the same reasoning, $(1 + u_1z) \mid a$, which together with $1 + u_1z = ac$ give $1 + u_1z = a$ and $c = 1$. $\square$

Remark. Note that the formula $u \in_0(y, z)$ is Σ_0 in $I\Sigma_1$.

1.49 Definition ($I\Sigma_1$). (1) (y, z) *o-codes* a sequence of length x if for each $u < x$ there is a $v < y$ such that $(u, v) \in_0(y, z)$. If this is the case then $(y, z)_u$ is the least $v < y$ such that $(u, v) \in_0(y, z)$.

(2) (y, z) is an *exponential sequence* of length x ($Exseq(y, z, x)$) if (y, z) o-codes a sequence of length x , $x \geq 1$, $(y, z)_0 = 1$ and, for each $u < x - 1$, $(y, z)_{u+1} = 2 * (y, z)_u$.

(3) $exp(x, v)$ if $(\exists y, z)(Exseq(y, z, x + 1) \& (y, z)_x = v)$.

1.50 Theorem. (1) $I\Sigma_1 \vdash (\forall x)(\exists! v)(exp(x, v))$.

(2) The formula $exp(x, v)$ is Δ_1 in $I\Sigma_1$.

(3) If we define in $I\Sigma_1$ the function 2^x by $exp(x, v)$ then all axioms of $I\Sigma_0(exp)$ are provable in $I\Sigma_1$.

Proof. (1) First show in $I\Sigma_1$ that if $Exseq(y, z, x)$, $Exseq(y', z', x')$ and $x \leq x'$ then, for each $u < x'$, $(y, z)_u = (y', z')_u$ (uniqueness). This is proved by induction on u , the formula in question being Σ_0 in $I\Sigma_1$. Similarly we prove that $Exseq(y, z, x)$ and $u < x$ implies $u < (y, z)_u$; furthermore, $Exseq(y, z, x)$ and $u < v < x$ implies $(y, z)_u < (y, z)_v$.

Then prove the following by Σ_1 induction on x :

$$x \geq 1 \rightarrow (\exists y, z) Exseq(y, z, x).$$

(Note that the formula $(\exists y, z)Exseq(y, z, x)$ is Σ_1 in $I\Sigma_1$ by contraction of quantifiers.) The assertion is evident for $x = 1$. Assume $x \geq 1$, $Exseq(y, z, x)$ and $(y, z)_{x-1} = q$; put $q' = (x, 2q)$. By 1.46, there are y', z' such that

$$(\forall u \leq q')(u \in_0 (y', z') \equiv (u \in_0 (y, z) \& (\exists i < x)(\exists v < u)(u = (i, v)) \vee u = (x, 2q))).$$

We get $Exseq(y', z', x + 1)$.

(2) Clearly, $exp(x, v)$ is Σ_1 as defined; but since we have (1), $\neg exp(x, v)$ is equivalent in $I\Sigma_1$ to $(\exists w)(x \neq v \& exp(x, w))$. Thus the result follows.

(3) We proceed in $I\Sigma_1$. Clearly, $2^0 = 1$. If $2^x = v$ and $2^{x+1} = w$ then we have $Exseq(y, z, x + 1)$, $Exseq(y', z', x + 2)$, $(y, z)_x = v$, $(y', z')_{x+1} = w$; but, by the claim in the proof of (1), $(y', z')_x = v$ and thus $w = 2v$. Thus the axioms for 2^x are provable. It remains to prove that induction for $\Sigma_0(exp)$ formulas is provable. But this follows from the fact that each $\Sigma_0(exp)$ formula is Δ_1 in $I\Sigma_1$. Let us prove this fact.

To prove that atomic $\Sigma_0(exp)$ formulas are Δ_1 it suffices to show that for each term t of $L_0(exp)$ and each variable x not occurring in t , the formula $t = x$ is Δ_1 in $I\Sigma_1$. This is clear for t atomic and the induction step for S , $+$ and $*$ is easy. (For example, $x = t + s$ is equivalent to $(\exists u, v \leq x)(u = t \& v = s \& x = u + v)$; if $u = t$ and $v = s$ are Δ_1 in $I\Sigma_1$ then so is $x = t + s$ by 1.45.) Consider $x = 2^t$ and let the formula $u = t$ be Δ_1 . It is sufficient to remember that $x = 2^u$ is Δ_1 in $I\Sigma_1$ (see (2) above); $x = 2^t$ is equivalent to $(\exists u \leq x)(x = 2^u \& u = t)$.

For the rest of the proof it suffices to recall that formulas Δ_1 in $I\Sigma_1$ are closed under connectives and bounded quantifiers. $\square$

*

1.51 Discussion and Definition. (1) Recall that we have Σ_n formulas and Π_n formulas; these are particular formulas of the language of arithmetic. Then we have Σ_n and Π_n sets of natural numbers; i.e. sets defined (in the standard model N by Σ_n formulas and Π_n formulas respectively). Δ_n sets are sets that are both Σ_n and Π_n (cf. Sect. 0). In 1.19 we defined a formula to be Σ_n (Π_n , Δ_n) in a theory T . Now we turn our attention to sets of natural numbers defined by such a formula. Instead of saying that a set X is defined by a formula that is Σ_n in T we say that X is T -provably Σ_n (similarly for Π_n , Δ_n). This generalizes to $X \subseteq N^k$, $k = 2, 3, \dots$

Clearly, if T is sound, i.e. N is a model of T and X is T -provably Σ_n (Δ_n etc.) then X is Σ_n (etc.). The converse need not be true, cf. Chap. IV, Sect. 3.

(2) A formula $\varphi(x, y)$ defines a total function in T if $T \vdash (\forall x)(\exists! y)\varphi(x, y)$. We may then extend T by defining a new function symbol F and the axiom $\varphi(x, F(x))$. We may again call the resulting theory T but care is necessary when dealing with hierarchies of formulas, e.g. we distinguish Σ_0 -formulas

and $\Sigma_0(exp)$ -formulas. If the formula φ defining F in T is Σ_n in T then we say that F is Σ_n in T , etc.

Clearly, if $\varphi(x, y)$ defines a total function in T and T is sound then φ defines a total function in N . A function $f : N \rightarrow N$ is *T-provably total* if it has a definition $\varphi(x, y)$ which defines a total function in T . The function f is *T-provably Σ_n* (etc.) if it has a definition which is Σ_n in T . (This is a particular case of (1).) The function f is a *T-provably total Σ_n function* if it has a definition φ which is Σ_n in T and defines a total function in T . In particular we call f *T-provably recursive* if it is *T-provably total Σ_1* ; since we shall often be interested in $I\Sigma_1$ -provably recursive functions we shall call them just *provably recursive*.

1.52 Lemma. Assume $T \supseteq I_{open}$. (1) If F is a function symbol Σ_1 in T then F is Δ_1 in T .

(2) If ψ is Δ_1 in T and F is a function symbol Δ_1 in T then the formula $(\exists x \leq F(y))\psi$ is Δ_1 in T .

Proof. (1) If $\varphi(x, y)$ is as above then the formula $(\exists y' \neq y)\varphi(x, y')$ is Σ_1 in T and defines the complement of F ; thus its negation is Π_1 and defines F .

(2) Let ψ' be $(\exists x \leq F(y))\varphi$. Clearly, ψ' is Σ_1 (by contraction of quantifiers). But in T , ψ' is also equivalent to $(\forall z)(z = f(y) \rightarrow (\forall x \leq z)\psi)$, which is Π_1 in T . $\square$

1.53 Lemma. If $T \supseteq I_{open}$ then *T-provably recursive* functions are closed under composition.

Proof. By contraction of quantifiers. $\square$

1.54 Theorem. If $T \supseteq I\Sigma_1$ then *T-provably recursive* functions are closed under primitive recursion. Thus each primitive recursive function is provably recursive.

For T containing $I\Sigma_1$ the converse is also true; thus provably recursive functions are exactly all primitive recursive functions. The converse inclusion will be proved in Chap. IV. The present theorem is an immediate consequence of the following lemma:

1.55. Let $T \supseteq I\Sigma_1$ and let $\varphi(\mathbf{x}, y)$ and $\psi(\mathbf{x}, u, v, y)$ be Σ_1 and define total functions in T , i.e. $T \vdash (\forall \mathbf{x})(\exists! y)\varphi(\mathbf{x}, y)$ and $T \vdash (\forall \mathbf{x}, u, v)(\exists! y)\psi(\mathbf{x}, u, v, y)$. Then there is a Σ_1 formula $\chi(\mathbf{x}, z, y)$ such that

$$T \vdash (\forall \mathbf{x}, z)(\exists! y)\chi(\mathbf{x}, z, y),$$

$$T \vdash (\forall \mathbf{x})(\forall y)(\chi(\mathbf{x}, 0, y) \equiv \varphi(\mathbf{x}, y),$$

$$T \vdash (\forall \mathbf{x}, z, y, y_1)(\chi(\mathbf{x}, z, y) \& \chi(\mathbf{x}, z + 1, y_1) \rightarrow \psi(\mathbf{x}, z, y, y_1)).$$

Thus if the functions defined by φ, ψ, χ are denoted by F, G, H respectively then $T \vdash F(\mathbf{x}, 0) = G(\mathbf{x})$ and $T \vdash F(\mathbf{x}, z + 1) = H(\mathbf{x}, z, f(\mathbf{x}, z))$.

Proof. χ just describes the course of values (*Exseq* was a particular case):

$$\chi(\mathbf{x}, z, y) \equiv (\exists s)(Seq(s) \& lh(s) = z + 1 \& (s)_0 = G(\mathbf{x}) \& (s)_x = y \& (\forall i < z)((s)_{i+1} = H(\mathbf{x}, i, (s)_i)).$$

Transcribing this with the help of φ, ψ is trivial but tiresome; clearly, χ is Σ_1 in T . $I\Sigma_1$ was used to prove $(\forall \mathbf{x}, z)(\exists y)\chi(\mathbf{x}, z, y)$; uniqueness is easy to prove and a pedantic elaboration of details of the proof of $F(\mathbf{x}, z + 1) = H(\mathbf{x}, z, F(\mathbf{x}, z))$ is left to the reader.

1.56 Remark. The lemma says (in contradistinction to 1.54) that inside $I\Sigma_1$ we may define total Δ_1 functions from other Δ_1 functions by primitive recursion. Note that this generalizes easily to primitive recursion on the course of values, cf. 0.44.

*

We now describe some concrete consequences of the preceding lemma.

1.57 Lemma. In $I\Sigma_1$ we may define total Δ_1 functions Σ and Π (sum and product of a sequence) such that $I\Sigma_1$ proves the following:

$$\begin{aligned} \Sigma x &= 0 \text{ if } x = 0 \text{ or } \neg Seq(x), \\ \Sigma(s \smallfrown \langle x \rangle) &= (\Sigma s) + x, \\ \Pi x &= 1 \text{ if } x = 0 \text{ or } \neg Seq(x), \\ \Pi(s \smallfrown \langle x \rangle) &= (\Pi s) * x. \end{aligned}$$

Proof. Left as an exercise. (Given s , prove by induction that for each $i \leq lh(s)$ there is a sequence s' of partial sums of length i such that $(s')_0 = (s)_0$ and, for $j < i - 1$, $(s')_{j+1} = (s')_j + (s)_{j+1}$. Put $\Sigma s = y$ if there is a sequence s' of partial sums of s of length $lh(s)$ such that $(s')_{lh(s)} = y$. Similarly for Π .) $\square$

1.58 Theorem. (1) In $I\Sigma_1$ we may Δ_1 define general power and factorial functions; i.e. total functions x^y and $x!$ such that the formulas $z = x^y$ and $z = x!$ are Δ_1 in $I\Sigma_1$ and $I\Sigma_1$ proves the following:

$$\begin{aligned} x^0 &= 1 \text{ and } x^{S(y)} = x^y * x, \\ 0! &= 1 \text{ and } (S(x))! = x! * S(x). \end{aligned}$$

(2) $I\Sigma_1$ proves that there are infinitely many primes. In $I\Sigma_1$ we may define an increasing Δ_1 enumeration of all primes.

(3) $I\Sigma_1$ proves the prime factorization theorem.

Proof. (1) follows directly from 1.55. To prove (2) work in $I\Sigma_1$ and take any x ; we show that there is a prime $p \geq x$. Let $z = \text{hull}(x)$, i.e. $(\forall u < x)(u \mid z)$ and take $z + 1$. By 1.26, there is a $p \mid (z + 1)$, but p is distinct from all $u < x$. (This is the classic Euclid's proof.) For each x , let $lp(x)$ be the least prime number greater than x ; by what we have just proved, lp is a total Δ_1 function. Thus the function

$$\begin{aligned} p_0 &= 2, \\ p_{x+1} &= lp(p_x) \end{aligned}$$

is Δ_1 and total – both provably in $I\Sigma_1$. This is the desired increasing enumeration of all primes.

(3) A sequence s is a *prime decomposition* if all members of s are primes and the sequence is non-decreasing, i.e. $(s)_i \leq (s)_{i+1}$ for all $i < lh(s) - 1$. We claim that for each $x > 0$ there is a unique prime decomposition s such that $\prod s = x$. Existence is proved by induction: the prime decomposition of 1 is the empty sequence $\emptyset$. Let $x > 1$ and assume $(\forall y < x)(y > 0 \rightarrow y \text{ has a prime decomposition})$. Let p be the largest prime dividing x (it exists by the least number principle for Σ_0 formulas) and take the y such that $x = p * y$ (divide x by p). Now $y < x$, so let s be a prime decomposition of y . Then $s \frown \langle p \rangle$ is a prime decomposition of x . $\square$

1.59 Remark. (1) Prove the uniqueness of the prime decomposition of x as an exercise.

(2) Many theorems of elementary number theory formalize easily in $I\Sigma_1$ together with their proofs; for example, the proof of *Bertrand's postulate* (saying that for each $x > 0$ there is a prime number p such that $x \leq p \leq 2x$) as given in [Hardy-Wright] can be easily rewritten in $I\Sigma_1$.

(3) Moreover, in Theorems 1.56, 1.57 $I\Sigma_1$ may be replaced by $I\Sigma_0(exp)$ but proofs then cost some additional effort since we do *not* have 1.54–1.55 for $I\Sigma_0(exp)$. Instead we have the following: $I\Sigma_0(exp)$ -provably total $\Sigma_0^{exp}(exp)$ functions are closed under bounded primitive recursion, i.e. if G, H, K are $I\Sigma_0(exp)$ -provably total $\Sigma_0^{exp}(exp)$ functions, F results from G, H by primitive recursion and F is provably majorized by K then F is $I\Sigma_0(exp)$ -provably total $\Sigma_0^{exp}(exp)$ function (and we have the corresponding lemma analogous to 1.55). The reader may elaborate details as an exercise.

(d) Arithmetization of Metamathematics: Partial Truth Definitions

Recall our investigations in 0.50–0.55 (beginning arithmetization of metamathematics): we showed there that various logical sets, functions, etc. are Δ_1 in N . As we promised there, we shall now strengthen these results and

develop them further; we are going to show that logical notions (like formulas, terms etc.) are Δ_1 in $I\Sigma_1$ and that $I\Sigma_1$ proves their basic properties. We shall detail careful formulations; proofs consist more or less in checking that informal proofs presented in Sect. 0 can be read as proofs in $I\Sigma_1$. Our gain will be twofold: We shall see that some reasonable parts of logic formalize in $I\Sigma_1$ and secondly, we shall be able to expand expressive possibilities of $I\Sigma_1$ by introducing variables for Σ_n (Π_n , Δ_n) sets of numbers. This will be very useful.

1.60 Theorem. Let $T \supseteq I\Sigma_1$, let $At^\bullet, Op^\bullet, Ar^\bullet$ be formulas Δ_1 in T and assume that T proves $At^\bullet, Op^\bullet$ to be disjoint, $At^\bullet$ non-empty and $Ar^\bullet$ to define a total function, i.e.

$$\begin{aligned} T \vdash (\forall x)(Op^\bullet(x) \rightarrow (\exists! y)Ar^\bullet(x, y)), \quad (\text{write } y = Ar^\bullet(x) \text{ for } Ar^\bullet(x, y)) \\ T \vdash (\exists x)(At^\bullet(x) \& (\forall y)(At^\bullet(y) \rightarrow \neg(Seq^\bullet(y) \& Op^\bullet((y)_0))) \\ \text{cf. (0.50).} \end{aligned}$$

Then there are formulas $Expr^\bullet, Appl^\bullet$ that are Δ_1 in T and such that T proves $(Expr^\bullet, Appl^\bullet)$ to be a free algebra of type $(Op^\bullet, Ar^\bullet)$ generated by $At^\bullet$, i.e.

$$\begin{aligned} T \vdash At^\bullet(x) \rightarrow Expr^\bullet(\langle x \rangle), \\ T \vdash Appl^\bullet(o, s, y) \\ \equiv (Op^\bullet(o), Seq^\bullet(s) \& lh(s) = Ar^\bullet(o), y = \langle o \rangle \frown Concseq(s)), \\ T \vdash Appl^\bullet(o, s, y) \& (\forall i < lh(s))(Expr^\bullet(((s)_i) \cdot) \rightarrow Expr^\bullet(y), \end{aligned}$$

and for each Σ_1 formula $\varphi(x)$ (possibly with parameters),

$$\begin{aligned} T \vdash (\forall x)(At^\bullet(x) \rightarrow \varphi(x)) \& \\ \& ((\forall o, s, y)(Appl^\bullet(o, s, y) \& (\forall i < lh(s))\varphi((s)_i) \cdot) \rightarrow \varphi(y)) \rightarrow \\ \rightarrow (\forall x)(Expr^\bullet(x) \rightarrow \varphi(x)). \end{aligned}$$

Thus atomic expressions are expressions; applying an operation to a sequence of expressions of the appropriate length gives an expressions; each non-atomic expression uniquely determines its components; and $Expr$ is the least Σ_1 set containing all atomic expressions and closed under application of operations.

Convention. We shall identify atomic expressions $\langle x \rangle$ with atoms x if there is no danger of misunderstanding. (This corresponds to the usual convention of omitting superfluous brackets.)

Proof. Define $Appl^*(o, s) = \langle o \rangle \frown Concseq(s)$; we define w to be a derivation formalizing the definition in 0.51; define

$$Expr^*(s) \equiv (\exists q)(q \text{ is a derivation and } s \text{ is its last element}).$$

The rest of the proof consists in checking the proof of 0.51. $\square$

Now we could define a Δ_1 presentation of terms and formulas of an arbitrary language; instead, we restrict ourselves to the language of *arithmetic* leaving the general case to the reader as an exercise.

1.61 Theorem. In $I\Sigma_1$ we can define constants $S^*, +^*, *^*, =^*, \leq^*, \bar{0}^*, \neg^*, \rightarrow^*, \forall^*, \Delta_1$ predicates $Var^*, Term^*, Atform^*, Form^*$ and Δ_1 functions $Applterm^*, Applform^*$ such that basic properties of terms and formulas are provable. More precisely, $I\Sigma_1$ proves the following:

(1) $(Term^*, Applterm^*)$ is a free algebra over variables * and the constant $\bar{0}^*$ as atoms with the operations S^* (unary), $+^*, *^*$ (binary);

(2) An atomic formula * consists of $=^*$ or $\leq^*$ together with two terms:

$$Atform^*(x) \equiv (\exists s, t \leq x)(Term^*(s) \& Term^*(t) \\ \& . x = \langle =^*, s, t \rangle \vee x = \langle \leq^*, s, t \rangle).$$

(3) $(Form^*, Applform^*)$ is a free algebra over atomic formulas * as atoms with the following operations: $\neg^*$ (unary), $\rightarrow^*$ (binary), and each variable * .

(4) There are infinitely many variables * :

$$(\forall x)(\exists y \geq x) Var^*(y).$$

(5) Terms * are disjoint from formulas * :

$$(\forall x) \neg (Term^*(x) \& Form^*(x)).$$

In a still more transparent way this may be formulated as follows:

Write

$$x +^* y \text{ instead of } Applterm^*(+^*, \langle x, y \rangle),$$

$$x *^* y \text{ instead of } Applterm^*(|^*, \langle x, y \rangle),$$

$$S^*(x) \text{ instead of } Applterm^*(S^*, \langle x \rangle),$$

$$t =^* s \text{ instead of } \langle =^*, t, s \rangle,$$

$$t \leq^* s \text{ instead of } \langle \leq^*, t, s \rangle,$$

$$\neg^*(x) \text{ instead of } Applform^*(\neg^*, \langle x \rangle),$$

$$x \rightarrow^* y \text{ instead of } Applform^*(\rightarrow^*, \langle x, y \rangle),$$

$$(\forall^* u)x \text{ instead of } Applform^*(\forall^*, \langle u, \langle x, y \rangle \rangle) \text{ where } Var^*(u)$$

Then $I\Sigma_1$ proves the following:

- * there are infinitely many variables^{*},
- * each variable^{*} is a term^{*},
- * if t, s are terms^{*} then $t +^* s, t *^* s, S^*(t)$ are terms^{*},
- * atomic formulas^{*} have the form $t =^* s$ or $t \leq^* s$ where t, s are terms^{*},
- * if x, y are formulas^{*} and u is a variable^{*} then $x \rightarrow^* y, \neg^* x, (\forall^* u)x$ are formulas^{*}.

Furthermore, for each Σ_1 formula φ , $I\Sigma_1$ proves the following:

If $\varphi(\bar{0}^*)$, $(\forall x)(Var^*(x) \rightarrow \varphi(x))$ and $(\forall t, s \text{ terms}^*)(\varphi(t) \& \varphi(s) \rightarrow \varphi(S^*(t), \varphi(t +^* s), \varphi(t *^* s))$ THEN $(\forall x)(Term^*(x) \rightarrow \varphi(x))$.

Similarly for formulas^{*}.

Proof. Choose concrete natural numbers $n_1, \dots, n_9$ such that $I\Sigma_1 \vdash \neg Seq(\bar{n}_i)$ and put, in $I\Sigma_1$, $S^* = \bar{n}_1, +^* = \bar{n}_2, \dots, \forall^* = \bar{n}_8, v^* = \bar{n}_9$ (auxiliary); define e.g. $Var^*(x) \equiv (\exists y \leq x)(x = \langle v^*, y \rangle)$. Then apply 1.60 twice: once for $\bar{0}^*$ and variables^{*} as atoms and $S^*, +^*, *^*$ as operations to get $Term^*$ and then for atomic formulas^{*} as atoms ($Atform^*$ as in (2)) and $\neg^*, \rightarrow^*$ and variables^{*} as operations. $\square$

1.62 Definition ($I\Sigma_1$). Define $vr^*(y) = \langle v^*, y \rangle$ (y -th variable), $nm^*(\bar{0}) = \bar{0}^*$, $nm^*(x + 1) = S^*(nm^*(x))$ (x -th numeral). Clearly, this defines total Δ_1 functions in $I\Sigma_1$. We often write $\dot{x}$ instead of $nm^*(x)$.

1.63 Theorem. (Construction of a Δ_1 function by induction on terms.) Let G, H_1, H_2, H_3 be total Δ_1 functions in $I\Sigma_1$. Then there is a total Δ_1 function in $I\Sigma_1$ such that $I\Sigma_1$ proves the following:

$$\begin{aligned}
 F(x) &= G(x) \text{ if } Var^*(x) \text{ or } x = \bar{0}^*, \\
 F(x +^* y) &= H_1(x, y) \\
 F(x *^* y) &= H_2(x, y) \\
 F(S^*(x)) &= H_3(x) \text{ for } x, y \text{ terms}^*, \\
 F(x) &= \bar{0} \text{ for } \neg Term^*(x).
 \end{aligned}$$

Proof. F may be constructed by recursion on the course of values (cf. 0.44 and 1.56). $\square$

1.64 Remark. (1) In $I\Sigma_1$ we may define a total Δ_1 function assigning to each term^{*} u the finite set of its variables^{*}:

$$\begin{aligned}
 Var_of^*(u) &= \{u\} \text{ if } Var^*(u), \\
 Var_of^*(u) &= \emptyset \text{ if } u = \bar{0}^*, \\
 Var_of(x +^* y) &= (Var_of^*(x)) \cup Var_of^*(y), \\
 Var_of(x *^* y) &= (Var_of^*(x)) \cup Var_of^*(y), \\
 Var_of^*(S^*(x)) &= Var_of^*(x).
 \end{aligned}$$

(2) A completely analogous theorem on the construction of a Δ_1 function by induction on formulas is now evident; for example, we may define the set of all its free variables^{*} of a formula^{*} as follows:

$$\begin{aligned} \text{Freevar}^*(u =^* v) &= \text{Var_of}^*(u) \cup \text{Var_of}^*(v), \\ \text{Freevar}^*(u \leq^* v) &= \text{similarly}, \\ \text{Freevar}^*(\neg^* x) &= \text{Freevar}^*(x), \\ \text{Freevar}^*(x \rightarrow^* y) &= \text{Freevar}^*(x) \cup \text{Freevar}^*(y), \\ \text{Freevar}^*((\forall^* w)x) &= \text{Freevar}^*(x) \setminus \{w\}. \end{aligned}$$

Furthermore, in $I\Sigma_1$ we may define total Δ_1 functions Subst^* (substitution), Val^* (evaluation of terms) such that $I\Sigma_1$ proves the following:

(3) (Substitution into terms.)

$$\begin{aligned} \text{Subst}^*(x, x, t) &= t \text{ if } \text{Var}^*(x) \& \text{Term}^*(t), \\ \text{Subst}^*(t_1 +^* t_2, x, t) &= \text{Subst}^*(t_1, x, t) +^* \text{Subst}^*(t_2, x, t), \end{aligned}$$

if t_1, t_2, t are terms^{*} and x is a variable^{*}; similarly for $*^*, S^*$.

(4) (Substitution into formulas.)

$$\begin{aligned} \text{Subst}^*(t_1 =^* t_2, x, t) &= (\text{Subst}^*(t_1, x, t) =^* \text{Subst}^*(t_2, x, t)), \\ \text{Subst}^*(\neg^* z, x, t) &= \neg^* \text{Subst}^*(z, x, t), \\ \text{Subst}^*(z_1 \rightarrow^* z_2, x, t) &= (\text{Subst}^*(z_1, x, t) \rightarrow^* \text{Subst}^*(z_2, x, t)), \\ \text{Subst}^*((\forall^* u)z, x, t) &= (\forall^* u)z \text{ if } u = x, \\ \text{Subst}^*((\forall^* u)z, x, t) &= (\forall^* u)\text{Subst}^*(z, x, t) \text{ if } u \neq x. \end{aligned}$$

(5) (Value of a term.) If t is a term^{*} and z is an evaluation of its variables^{*}, i.e. a finite mapping whose domain consists of some variables^{*}, among them all variables^{*} of t then:

$$\begin{aligned} \text{Val}^*(t, z) &= z(t) \text{ if } t \text{ is a variable}^*, \\ \text{Val}^*(t_1 +^* t_2, z) &= \text{Val}^*(t_1, z) + \text{Val}^*(t_2, z) \end{aligned}$$

similarly for $*^*, S^*$ and $*, S$.

1.65 Remark. We have constructed a definition of sequences, terms, formulas, etc. that are Δ_1 in $I\Sigma_1$ and such that basic properties (in particular, closure properties) are provable in $I\Sigma_1$. Our definitions define the corresponding notions (sets, functions, etc.) in N ; for example, φ is a formula iff $N \models \text{Form}^*[\varphi]$ (this was discussed in Sect. 0). But since our definitions are Δ_1 in $I\Sigma_1$ we can use Σ_1 -completeness of $I\Sigma_1$ (see 1.9) to get the following:

(i) For each $t \in N$,

t is a term iff $N \models \text{Term}^*(\bar{t})$ iff $I\Sigma_1 \vdash \text{Term}^*(\bar{t})$,

t is $t_1 + t_2$ iff $N \models \bar{t} = (\bar{t}_1 +^\bullet \bar{t}_2)$ iff $I\Sigma_1 \vdash \bar{t} = \bar{t}_1 +^\bullet \bar{t}_2$,
similarly for $S, *$.

(ii) For each $\varphi \in N$,

φ is a formula iff $N \models \text{Form}^\bullet(\bar{\varphi})$ iff $I\Sigma_1 \vdash \text{Form}^\bullet(\bar{\varphi})$,

φ is $t = s$ iff $N \models \bar{\varphi} = (\bar{t} =^\bullet \bar{s})$ iff $I\Sigma_1 \vdash \bar{\varphi} = (\bar{t} =^\bullet \bar{s})$,

φ is $\varphi_1 \rightarrow \varphi_2$ iff $N \models \bar{\varphi} = (\bar{\varphi}_1 \rightarrow^\bullet \bar{\varphi}_2)$ iff $I\Sigma_1 \vdash \bar{\varphi} = (\bar{\varphi}_1 \rightarrow^\bullet \bar{\varphi}_2)$,

similarly for $t \leq s, \neg, (\forall y)$.

(iii) If t, s are terms, φ is a formula and x is a variable then

$$\begin{aligned} I\Sigma_1 \vdash \text{Subst}^\bullet(\bar{t}, \bar{x}, \bar{s}) &= \overline{\text{Subst}(t, x, s)}, \\ I\Sigma_1 \vdash \text{Subst}^\bullet(\bar{\varphi}, \bar{x}, \bar{s}) &= \overline{\text{Subst}(\varphi, x, s)}. \end{aligned}$$

1.66 Lemma. If $t(x_0, \dots, x_n)$ is a term whose variables are among $x_0, \dots, x_n$ then

$$I\Sigma_1 \vdash t(x_0, \dots, x_n) = \text{Val}^\bullet(\bar{t}, z(x_0, \dots, x_n)),$$

where $z(x_0, \dots, x_n)$ is the finite mapping associating with each variable $^\bullet \bar{x}_i$ the number x_i ($i = 0, \dots, n$).

Proof. Construct the corresponding $I\Sigma_1$ -proofs by induction over subterms of t : if t is x_i then $I\Sigma_1 \vdash x_i = x_i$; if t is $t_1 + t_2$ and $I\Sigma_1 \vdash t_i = \text{Val}^\bullet(\bar{t}_i, z)$ ($i = 1, 2$) then $I\Sigma_1 \vdash t_1 + t_2 = \text{Val}^\bullet(\bar{t}_1 + \bar{t}_2, z)$, etc. $\square$

1.67 Remark. Note that this may also be expressed as follows: write $v(w/u)$ for $\text{Subst}^\bullet(v, u, w)$ and $v(w_1/u_1, \dots, w_n/u_n)$ for iterated substitution. Then

$$I\Sigma_1 \vdash t(x_0, \dots, x_n) = \text{Val}(\bar{t}(\dot{x}_0/\bar{x}_0, \dots, \dot{x}_n/\bar{x}_n, \emptyset).$$

In particular, for t without variables we get $I\Sigma_1 \vdash t = \text{Val}(\bar{t})$, e.g. $\bar{3} + \bar{7} = \text{Val}(\bar{3} +^\bullet \bar{7}) = \bar{10}$.

1.68 Lemma. There is a formula $\Sigma_0^\bullet(x)$ (saying: x is a $\Sigma_0^\bullet$ -formula $^\bullet$) such that

- (1) $\Sigma_0^\bullet(x)$ is a Δ_1 formula in $I\Sigma_1$ and
- (2) $I\Sigma_1$ proves the following:
 - (i) Each atomic formula $^\bullet$ is $\Sigma_0^\bullet$,
 - (ii) $\Sigma_0^\bullet$ formulas $^\bullet$ are closed under connectives $^\bullet$ and [bounded quantifiers] $^\bullet$;
- (3) Furthermore, for each Σ_1 formula $\varphi(x)$, $I\Sigma_1$ proves the following:
 - (iii) If each atomic formula $^\bullet$ satisfies φ and formulas $^\bullet$ satisfying φ are closed under connectives $^\bullet$ and [bounded quantifiers] $^\bullet$ then each $\Sigma_0^\bullet$ formula $^\bullet$ satisfies φ .

Proof. Formalization of bounded quantifiers is clear: $(\forall^\bullet u \leq^\bullet v)z$ is just $(\forall^\bullet u)(u \leq^\bullet v \rightarrow^\bullet z)$. The proof is completely analogous to the proof of 1.60 (Δ_1 definition of expressions). $\square$

1.69 Lemma. There are formulas $\Sigma_y^\bullet(x)$, $\Pi_y^\bullet(x)$ with two free variables (read: x is a $\Sigma_y^\bullet$ formula^{*}, similarly for Π) such that (1) both $\Sigma_y^\bullet(x)$ and $\Pi_y^\bullet(x)$ are Δ_1 in $I\Sigma_1$, and (2) $I\Sigma_1$ proves the following:

- (i) for $y = 0$, $\Sigma_0^\bullet(x) \equiv \Pi_0^\bullet(x)$;
- (ii) $\Pi_{y+1}^\bullet(x)$ iff there is a variable^{*} $u \leq x$ and a $\Sigma_y^\bullet$ formula z such that $x = (\forall^\bullet u)z$;
- (iii) similarly for $\Sigma_{y+1}^\bullet$.

Proof. Exercise. □

We are now ready for a definition of satisfaction for Σ_0 formulas.

1.70 Theorem. There is a formula $Sat_0(z, e)$ which is Δ_1 in $I\Sigma_1$ and such that $I\Sigma_1$ proves Tarski's satisfaction conditions (cf. 0.6) for Σ_0 formulas^{*}, i.e. $I\Sigma_1$ proves the following:

- (i) $Sat_0(z, e) \rightarrow z$ is a $\Sigma_0^\bullet$ formula^{*} and e is an evaluation^{*} for z ,
- (ii) if z is $\Sigma_0^\bullet$ and $z = (u =^\bullet v)$ then

$$Sat_0(z, e) \equiv Val^\bullet(u, e) = Val^\bullet(v, e)$$

and similarly for $z = (u \leq^\bullet v)$;

- (iii) if z is $\Sigma_0^\bullet$ and $z = (\neg^\bullet u)$ then $Sat_0(z, e)$ iff $\neg Sat_0(u, e)$ and similarly for $z = (u \rightarrow^\bullet v)$;
- (iv) if z is $\Sigma_0^\bullet$ and $z = ((\forall^\bullet w_1 \leq^\bullet w_2)u)$ then $Sat_0(z, e)$ iff for each e' evaluation of u coinciding with e on $Freevar^\bullet(z) \setminus \{w_1\}$ and such that $e'(w_1)$ is defined and $e'(w_1) \leq e'(w_2)$ we have $Sat_0(u, e')$.

The proof is in 1.71–1.73.

1.71 Definition ($I\Sigma_1$). (1) q is a *partial satisfaction* for $\Sigma_0^\bullet$ formulas^{*} $\leq p$ and their evaluations^{*} by numbers $\leq r$ (in symbols: $PSat_0(q, p, r)$) if q is a finite mapping whose domain consists of all pairs (z, e) where z is $\Sigma_0^\bullet$, $z \leq p$, e is an evaluation^{*} for z , $e \subseteq (\leq p) \times (\leq r)$, $range(q) \subseteq \{0, 1\}$ and Tarski's conditions hold for q whenever those things in question are defined, i.e. for each $(z, e) \in dom(q)$,

- (ii) if $z = (u =^\bullet v)$ then $q(z, e) = \bar{1}$ iff $Val^\bullet(u, e) = Val^\bullet(v, e)$, similarly for $z = (u \leq^\bullet v)$;
- (iii) if $z = (\neg^\bullet u)$ then $q(z, e) = \bar{1}$ iff $q(u, e) = \bar{0}$, similarly for $z = (u \rightarrow^\bullet v)$;
- (iv) if $z = ((\forall^\bullet w_1 \leq^\bullet w_2)u)$ then $q(z, e) = \bar{1}$ iff for each $e' \subseteq (\leq p) \times (\leq r)$ as in (iv) above we have $q((u, e')) = \bar{1}$.

(Note that e is assumed to be defined only for (some) variables^{*} y such that $y \leq p$ and evaluates them by numbers $\leq r$.)

- (2) $Sat_0(z, e)$ iff there are q, p, r such that $PSat_0(q, p, r)$ and $q(z, e) = 1$.

1.72 Lemma. (1) $PSat_0$ is Δ_1 in $I\Sigma_1$.

(2) $I\Sigma_1$ proves that if q_1, q_2 are partial satisfactions for $\Sigma_0^\bullet$ then they coincide at the intersection of their domains.

(3) $I\Sigma_1$ proves that for each p, r , there is a q such that $PSat_0(q, p, r)$.

Proof. (1) Recall that Δ_1 includes $\Sigma_0^{exp}(exp)$ and also recall 1.52(2). For example, $dom(q)$ is characterized as follows:

$$\begin{aligned} &(\forall x \in dom(q))(\exists z, e \leq x)(x = (z, e) \& e \text{ is an evaluation}^\bullet \\ &\quad \text{for } z \& z \leq p \& e \leq r) \& \\ &(\forall z, e \leq H(p, r))(e \text{ is an evaluation}^\bullet \text{ for } z \& \\ &\quad \& z \leq p \& e \leq r. \rightarrow (z, e) \in dom(q)) \end{aligned}$$

where $H(p, r)$ is a term majorizing all such (z, e) ; take e.g. $(p + r + 2)^2$. The rest is left as an exercise.

(2) In $I\Sigma_1$ we prove the formula

$$\begin{aligned} &PSat_0(q_1, p_1, r_1) \& PSat_0(q_2, p_2, r_2) \rightarrow \\ &\rightarrow (\forall e < q_1)(\text{both } q_1(z, e) \text{ and } q_2(z, e) \text{ defined} \rightarrow q_1(z, e) = q_2(z, e)) \end{aligned}$$

by induction on z (in the form 1.45(3)): if $q_1(z, e)$ and $q_2(z, e)$ are defined and z is atomic then the conclusion follows by the definition of $PSat_0$; if z is $\neg^\bullet u$ are $u \rightarrow^\bullet v$ then by the inductive assumption, $q_1(u, e)$ and $q_2(u, e)$ are defined and equal (and the same for v); thus $q_1(z, e) = q_2(z, e)$. If z is $(\forall^\bullet w_1 \leq^\bullet w_2)u$ and $q_i(z, e)$ is defined then $e(w_2)$ is defined and $e(w_2) \leq r$; thus if $e' \subseteq (\leq p) \times (\leq r)$ is a finite mapping coinciding with e on all free variables $^\bullet$ of u except possibly w_1 , assigning to w_1 a value $\leq e(w_2)$ and undefined elsewhere then $q_i(u, e')$ is defined, and by the induction assumption, $q_1(u, e') = q_2(u, e')$. Thus $q_1(z, e) = q_2(z, e)$.

(3) We prove in $I\Sigma_1$ the formula

$$(*) \quad (\exists q)PSat_0(q, p, r)$$

by induction on p , with r a parameter. Nothing has to be proved for $p = \bar{0}$; assume $(*)$, let $PSat_0((q, p, r)$ and take $u = p + 1$. If u is neither a $\Sigma_0^\bullet$ formula nor a variable $^\bullet$ then we have $PSat_0(q, u, r)$; if u is a variable $^\bullet$ then we have to extend q to a q' such that q' is defined for all pairs (z, e) where z is a $\Sigma_0^\bullet$ formula $^\bullet$, $z \leq p$ and e is an evaluation for z , $e \subseteq (\leq u) \times (\leq r)$. Therefore e may be defined for u but the value is irrelevant for $q'(z, e)$ since u cannot occur in z . Thus put $q'(z, e) = q(z, e \downarrow)$ where $e \downarrow$ means the restriction of e to arguments different from u . Show using comprehension that q' exists; $PSat_0(q', u, r)$ is evident.

If u is a $\Sigma_0^\bullet$ formula then we have to discuss several cases; but note that now u is not a variable $^\bullet$ so we have only to investigate only evaluations

$e \subseteq (\leq p) \times (\leq r)$. We need to extend q to a q' such that $q'(u, e)$ is defined for all e just mentioned.

(i) u is atomic, $u = (t =^{\bullet} s)$: define

$$\begin{aligned} q'(u, e) &= \bar{1} \text{ iff } Val^{\bullet}(t) = Val^{\bullet}(s), \\ q'(u, e) &= \bar{0} \text{ otherwise,} \\ q'(z, e) &= q(z, e) \text{ if defined.} \end{aligned}$$

q' exists by comprehension.

(ii) u is $\neg^{\bullet} v$; since $v \leq p$, $q(v, e)$ is defined for all e in question. We put $q'(u, e) = \bar{1} - q(v, e)$, $q'(z, e) = q(z, e)$ if defined. Similarly for $u = (v \rightarrow^{\bullet} w)$.

(iii) u is $(\forall^{\bullet} w_1 \leq^{\bullet} w_2) v$. If $e \subseteq (\leq p) \times (\leq r)$ is an evaluation for u and $a = e(w_2)$ then extend q by defining $q'(u, e) = \bar{1}$ iff for all $e' \subseteq (\leq p) \times (\leq r)$ evaluations for v coinciding with e on $Freevar^{\bullet}(v) \setminus \{w_1\}$, defined also for w_1 and assigning to w_1 a value $e'(w_1) \leq a$ we have $q(v, e') = \bar{1}$; otherwise put $q'(u, e) = \bar{0}$. (Furthermore, $q'(z, e) = q(z, e)$ if defined.) Again q' exists by comprehension. This completes the proof. $\square$

1.73 Lemma ($I\Sigma_1$). (1) Sat_0 is Δ_1 in $I\Sigma_1$. (2) $I\Sigma_1$ proves Tarski's satisfaction conditions for $\Sigma_0^{\bullet}$ formulas^{*} and Sat_0 .

Proof. (1) As it stands, Sat_0 is clearly Σ_1 . But by the preceding lemma, $I\Sigma_1$ proves (assuming that z is a formula^{*} and e is an evaluation^{*}) $Sat_0(z, e)$ to be equivalent to

$$(\forall q, p, r)(PSat_0(q, p, r) \& q(z, e) \text{ defined} \rightarrow q(z, e) = 1).$$

Thus Sat_0 is Π_1 in $I\Sigma_1$.

(2) The only thing to check is the condition for bounded quantifiers. Assume $z = (\forall^{\bullet} w_1 \leq^{\bullet} w_2) u$ and $Sat_0(z, e)$; thus assume $PSat(p, q, r)$ and $q(z, e) = 1$. Then $z \leq p$ and for each $e' \subseteq (\leq p) \times (\leq r)$ such that e' coincides with e on $Freevar^{\bullet}(u) \setminus \{w_1\}$ and $e'(w_2) \leq e(w_1)$ we have $q(u, e') = 1$, i.e. $Sat_0(u, e')$. We have to get rid of the condition $e' \subseteq (\leq p) \times (\leq r)$. But if e' is as above except for the last condition then the restriction e'' of e to $Freevar^{\bullet}(u) \cup \{w_1\}$ does satisfy the condition $e'' \subseteq (\leq p) \times (\leq r)$ and for each q', p', r' such that $PSat(q', p', r')$ and both $q'(u, e')$ and $q'(u, e'')$ are defined, we have $q'(u, e') = q'(u, e'')$ (since ee', e'' coincide on relevant variables^{*}). Thus we have $Sat_0(u, e'')$ and $Sat_0(u, e')$. We have proved the implication to in 1.70(iv); the converse implication is easy. This completes the proof of the lemma and of 1.70. $\square$

1.74 Definition. For each $n > 0$ we define in $I\Sigma_1$ predicates $Sat_{\Sigma, n}$ and $Sat_{\Pi, n}$ as follows:

$$Sat_{\Sigma, 0}(z, e) \equiv Sat_{\Pi, 0}(z, e) \equiv Sat_0(z, e);$$

given $Sat_{\Sigma,n}$ we define

$$Sat_{\Pi,n+1}(z, e) \equiv .[z \text{ has the form } (\forall^* x)u \text{ where } u \text{ is } \Sigma_n^* \\ e \text{ evaluates free variables}^* \text{ of } z \text{ and for each} \\ \text{evaluation } e' \text{ for } u \text{ coinciding with } e \\ \text{on } Freevar^*(u) \setminus \{x\} \text{ we have } Sat_{\Sigma,n}(u, e')].$$

Similarly for $Sat_{\Sigma,n+1}$ (from $Sat_{\Pi,n}$).

1.75 Theorem. (1) For each $n \leq 1$, $Sat_{\Sigma,n}$ is Σ_n in $I\Sigma_1$ and $Sat_{\Pi,n}$ is Π_n in $I\Sigma_1$.

(2) $I\Sigma_1$ proves Tarski's satisfaction conditions for $Sat_{\Sigma,n}$ and Σ_n^* formulas * as well as for $Sat_{\Pi,n}$ and Π_n^* formulas * ; i.e. it proves analogs of 1.70(i)–(iv) and, in addition,

(v) if $m \leq n$, z is Σ_m^* and $z = (\exists^* x)u$ then $Sat_{\Sigma,n}(z, e)$ iff there is an evaluation * e' of u coinciding with e on $Freevar^*(z)$ and such that $Sat_{\Sigma,n}(u, e')$;

(v') if $m \leq n$, z is Π_m^* and $z = (\forall^* x)u$ then $Sat_{\Sigma,n}(z, e)$ iff for all evaluations e' of u coinciding with e on $Freevar^*(z)$ we have $Sat_{\Sigma,n}(u, e')$;

(v''), (v'') similarly for $Sat_{\Pi,n}$.

Proof. (1) is obvious from 1.70 by induction on n . Also (2) is easy to prove from the definitions. $\square$

1.76 Corollary ("It's snowing"-It's Snowing-Lemma). If $\varphi(x_0, \dots, x_n)$ is Σ_n then

$$I\Sigma_1 \vdash \varphi(x_0, \dots, x_n) \equiv Sat_{\Sigma,n}(\overline{\varphi}(\dot{x}_0/\overline{x_0}, \dots, \dot{x}_n/\overline{x_n}, \emptyset).$$

This formalizes Tarski's example: the sentence "it's snowing" is true iff it's snowing. Recall that $\overline{\varphi}(\dot{x}_0/\overline{x_0}, \dots)$ is an abbreviation of

$$Subst^* \dots Subst^*(Subst^*(\overline{\varphi}, vr^*(\overline{0}), \dot{x}_0), vr^*(\overline{1}), \dot{x}_1), \dots, vr^*(\overline{n}), \dot{x}_n).$$

1.77 Remark. Observe that 1.70 is just a theorem in $I\Sigma_1$; to prove it we needed only a finite fragment of $I\Sigma_1$ (since each proof of a single theorem uses finitely many axioms). Similarly, for each fixed $n \leq 1$, Tarski's satisfaction conditions for Σ_n^* (or Π_n^*) are expressed by finitely many sentences (or just one - take conjunction). Thus only a finite fragment of $I\Sigma_1$ is needed to prove them. This fact will be used in the next section to show finite axiomatizability of some fragments of arithmetic (including $I\Sigma_1$).

1.78 Definition. We shall close this long section by extending the expressive possibilities of theories containing $I\Sigma_1$ by introducing variables for some definable possibly infinite sets of numbers. Let n be a fixed natural number. We make in $I\Sigma_1$ the following definitions:

- (1) c is a (code of a) $\Sigma_n^\bullet$ set^{*} iff c is a $\Sigma_n^\bullet$ formula^{*} and its only free variable^{*} is $vr^\bullet(0)$ (the 0-th variable^{*}).
- (2) $x \in \Sigma_{\mathcal{L},n} c$ iff c is a $\Sigma_n^\bullet$ set^{*} and $Sat_{\Sigma,n}(c, [x])$ (where $[x]$ is the evaluation assigning x to $vr^\bullet(0)$).
- (3) $\Pi_n^\bullet$ sets and $x \in \Pi_{\mathcal{L},n} c$ are defined dually.

Note that the formula $x \in \Sigma_{\mathcal{L},0} c$ is Δ_1 in $I\Sigma_1$ and, for $n \geq 1$, the formulas $x \in \Sigma_{\mathcal{L},n} c$, $x \in \Pi_{\mathcal{L},n} c$ are Σ_n and Π_n in $I\Sigma_1$ respectively. Given n we may introduce new variables in $I\Sigma_1$ ranging over $\Sigma_n^\bullet$ sets. If n is clear from the context we may use just $X, Y, \dots$. The following lemma serves as an example of its usage.

1.79 Lemma ($I\Sigma_1$). For each $\Sigma_1^\bullet$ total function^{*} F there is a $\Sigma_1^\bullet$ total function^{*} G such that $G(0) = 0$ and, for each x , $G(x+1) = F(G(x))$.

Proof. Very similar to the proof of 1.54, with the only difference being that now we have *one* proof in $I\Sigma_1$ for all total $\Sigma_1^\bullet$ functions^{*}, not a schema of theorems with infinitely many proofs. $\square$

1.80 Remark. Note that if φ is a Σ_n formula whose only free variable is x_0 (the 0-th variable) then $I\Sigma_1$ proves that $\bar{\varphi}$ is a $\Sigma_n^\bullet$ set and that

$$x \in \Sigma_{\mathcal{L},n} \bar{\varphi} \equiv Sat_{\Sigma,n}(\bar{\varphi}, [x]) \equiv \varphi(x).$$

(cf. “it’s snowing”-it’s snowing Lemma 1.76).

1.81 Definitions Continued. Let $n \geq 0$ be fixed. In $I\Sigma_1$ we define a $\Delta_n^\bullet$ set to be a pair (c, d) where c is a $\Sigma_n^\bullet$ set, d is a $\Pi_n^\bullet$ set and $(\forall x)(x \in \Sigma_{\mathcal{L},n} c \equiv x \in \Pi_{\mathcal{L},n} d)$. Observe that the formula saying that x is a $\Delta_n^\bullet$ set is Π_{n+1} in $I\Sigma_1$. We may define $x \in \Delta_{\mathcal{L},n} (c, d)$ in the obvious way; again, this formula is Π_{n+1} in $I\Sigma_1$. But observe that $I\Sigma_1$ proves the following:

$$e = (c, d) \ \& \ \Delta_n^\bullet\text{-set}(e) \rightarrow (\forall x)(x \in \Delta_{\mathcal{L},n} e \equiv x \in \Sigma_{\mathcal{L},n} c \equiv x \in \Pi_{\mathcal{L},n} d).$$

Furthermore, $x \in \Sigma_{\mathcal{L},n} c$ is Σ_n and $x \in \Pi_{\mathcal{L},n} d$ is Π_n in $I\Sigma_1$.

1.82 Remark. Let us *summarize* what we have done in the present section. We first introduced Robinson’s arithmetic Q and proved it to be Σ_1 -complete. We further introduced I_{open} and proved in it some high-school laws for numbers: associativity and commutativity of addition and multiplication, distributivity, cancellation, monotonicity, etc. Furthermore, we exhibited the pairing function in I_{open} . Then we showed in $I\Sigma_0$ some properties of divisibility (and left thorough investigation of $I\Sigma_0$ to Chap. V). In $I\Sigma_0(\exp)$ we developed a coding of finite sequences and finite sets and proved some basic facts

about finite sets and their cardinalities. Then we turned to IS_1 ; we showed that IS_1 -provably recursive functions are closed under primitive recursion and then we developed arithmetization of metamathematics in this theory. We defined terms^{*} and formulas^{*} and proved their basic properties. Our final development has been a definition of partial satisfactions; for each n we have the formulas $Sat_{\Sigma,n}$ and $Sat_{\Pi,n}$ with provable Tarski's properties. This enables us to introduce variables for Σ_n^* sets, etc. (for any fixed n). This basic apparatus will be used throughout Chaps. I-IV.

2. Fragments of First-Order Arithmetic

Recall that in Sect. 1 we already investigated some fragments of first-order arithmetic, notably $IS_0(exp)$ and IS_1 . Now we are going to investigate systematically fragments obtained by postulating a number-theoretic principle as a scheme for all formulas of a certain class. In subsection (a) we shall deal with fragments based on induction, the least number principle and collection; in subsection (b) we shall study various other principles. Recall also that in Sect. 1 we exhibited satisfaction for Σ_n -formulas (Π_n -formulas) for any fixed n . In subsection (c) we shall use this device to show that most of our fragments are finitely axiomatizable; then we shall generalize and show that under some assumptions we can exhibit in IS_1 a reasonable satisfaction for the relativized arithmetical hierarchy; namely for formulas Σ_n in a set X . In subsection (d) we apply this to particular fragments; this will give us techniques very useful in the following section. Subsection (e) is an appendix presenting an alternative approach to fragments in the logic without function symbols. Results of this section will be used throughout the book.

(a) Induction and Collection

2.1. Here we shall investigate the following four axioms that we met already in Sect. 1:

- $$\begin{aligned}
 (I\varphi) \quad & \varphi(0) \ \& \ (\forall x)(\varphi(x) \rightarrow \varphi(S(x)) \rightarrow (\forall x)\varphi(x) \\
 (I'\varphi) \quad & (\forall x)[(\forall y < x)\varphi(y) \rightarrow \varphi(x)] \rightarrow (\forall x)\varphi(x) \\
 (L\varphi) \quad & (\exists x)\varphi(x) \rightarrow (\exists x)(\varphi(x) \ \& \ (\forall y < x)\neg\varphi(y)) \\
 (B\varphi) \quad & (\forall u)[(\forall x \leq u)(\exists y)\varphi(x, y) \rightarrow (\exists v)(\forall x \leq u)(\exists y \leq v)\varphi(x, y)]
 \end{aligned}$$

They are called the *successor induction axiom* given by φ , the *order induction axiom* by φ , the *least number axiom* given by φ and the *collection axiom* given by φ respectively.

Both the axioms of induction and the least number principle appear to be sufficiently clear. Let us give a verbal formulation of collection: think of

$\varphi(x, y)$ as defining a multi-valued function Φ : $\varphi(x, y)$ says that y is a possible value of Φ for x . Call Φ *total* beneath u if each $x \leq u$ has at least one possible value. Call Φ *cofinal* on u if for each v there is an $x \leq u$ such that all possible Φ -values of x are bigger than v . B_φ says: if Φ is total beneath u then it is not cofinal on u .

2.2 Some Classes of Formulas. Recall Σ_n and Π_n formulas. If $\varphi(x_0 \dots x_k)$ is any formula then an *instance* of φ is $\varphi(t_1, \dots, t_k)$ where t_i are terms, t_i free for x_i in φ . $\Sigma_0(\varphi)$ is the smallest class of formulas that (1) contains all atomic formulas and all instances of φ and (2) is closed under connectives and bounded quantification. $\Sigma_n(\varphi)$ formulas are defined from $\Sigma_0(\varphi)$ formulas in the same manner as Σ_n was defined from Σ_0 ; thus a $\Sigma_n(\varphi)$ formula consists of a block of n alternating quantifiers, the first being $\exists$, followed by a $\Sigma_0(\varphi)$ formula. $\Sigma_m(\Sigma_n)$ is the union of all $\Sigma_m(\varphi)$, φ being Σ_n . Similarly for $\Pi_m(\Sigma_n)$ etc.

A formula ψ is $\Sigma_n(\varphi)$ in a theory T if there is a $\Sigma_n(\varphi)$ formula χ such that $T \vdash \psi \equiv \chi$. A formula ψ is $\Delta_n(\varphi)$ in T if it is both $\Sigma_n(\varphi)$ in T and $\Pi_n(\varphi)$ in T .

Observe that for each $\varphi(x_0 \dots x_k)$ there is a formula $\psi(u)$ with exactly one free variable such that each $\Sigma_0(\varphi)$ formula is $\Sigma_0(\psi)$ in T and vice versa (we assume $T \supseteq I\Sigma_0(exp)$ for simplicity): take $\varphi(u)$ to be

$$(\exists x_0 \leq u) \dots (\exists x_k \leq u)(u = \langle x_0, \dots, x_k \rangle \& \varphi(x_0, \dots, x_k)).$$

2.3 Theories. Recall Q . For some technical reasons (see below) introduce also a theory Q' defined as Q plus the axiom $x < Sx$ (Note $I_{open} \vdash x \leq Sx$.) We shall study the following theories:

$$\begin{aligned} I\Sigma_n &= Q \cup \{I\varphi \mid \varphi \in \Sigma_n\}, \\ I'\Sigma_n &= Q' \cup \{I'\varphi \mid \varphi \in \Sigma_n\}, \\ L\Sigma_n &= Q' \cup \{L\varphi \mid \varphi \in \Sigma_n\}. \end{aligned}$$

Similarly for $I\Gamma$, $I'\Gamma$, $L\Gamma$, where Γ is Π_n , $\Sigma_0(\Sigma_n)$ etc.

(Caution: $I\Delta_n$ has not any meaning yet, see below!) Further we define

$$B\Gamma = I\Sigma_0 \cup \{B\varphi \mid \varphi \in \Gamma\}$$

for Γ as above, i.e. we have $B\Sigma_n$, $B\Pi_n$ etc. Note that by definition $B\Gamma$ contains $I\Sigma_0$.

$I\Delta_n$ is Q together with the scheme

$$(\forall x)(\sigma(x) \equiv \pi(x)) \rightarrow I\sigma$$

for each pair $\sigma \in \Sigma_n$, $\pi \in \Pi_n$ (thus the axiom says: if σ and π are equivalent then induction holds for σ – and obviously also for Π). Similarly $L\Delta_n$.

In 2.4–2.5 and 2.7 we formulate the principal facts about these theories.

2.4 Theorem. For each n , the following nine theories are mutually equivalent:

$$\begin{array}{lll} I\Sigma_n, & I\Pi_n, & I\Sigma_0(\Sigma_n), \\ I'\Sigma_n, & I'\Pi_n, & I'\Sigma_0(\Sigma_n), \\ L\Sigma_n, & L\Pi_n, & L\Sigma_0(\Sigma_n). \end{array}$$

2.5 Theorem. For each n ,

- (1) $I\Sigma_{n+1} \Rightarrow B\Sigma_{n+1} \Rightarrow I\Sigma_n$;
 - (2) $B\Sigma_{n+1} \Leftrightarrow B\Pi_n \Leftrightarrow L\Delta_{n+1} \Rightarrow I\Delta_{n+1}$,
 - (3) For each Σ_n -formula φ , the formula $(\forall x \leq y)\varphi$ is Σ_n in $B\Sigma_n$.
- (Here $\Rightarrow$ means “contains” (i.e. “proves all axioms of”) and $\Leftrightarrow$ means “is equivalent to”).

Remark. It is unknown whether $I\Delta_n$ and $L\Delta_n$ are equivalent; all the other arrows will be shown to be strict in Chap. IV (the theories in question are not equivalent).

2.6 Definition. Let $\varphi(x_0, \dots, x_k)$ be a formula. Define in $I\Sigma_1$: q is a z -piece of φ if q is a finite mapping, $\text{dom}(q) = (< z)^{k+1}$, $\text{range}(q) \subseteq \{0, 1\}$ and

$$(\forall x_0 < z) \dots (\forall x_k < z)(q(\langle x_0, \dots, x_k \rangle) = 1 \equiv \varphi(x_1 \dots x_k)).$$

φ is *piecewise coded* in $T \supseteq I\Sigma_1$ if T proves $(\forall z)(\exists q)(q \text{ is a } z\text{-piece of } \varphi)$.

We often write “p.c.” for “piecewise coded”.

2.7 Theorem. For each $n \geq 1$,

- (1) each $\Sigma_0(\Sigma_n)$ formula φ is p.c. in $I\Sigma_n$;
- (2) if φ is Δ_{n+1} in $B\Sigma_{n+1}$ then φ is p.c. in $B\Sigma_{n+1}$.

In particular, each Σ_n formula is p.c. in $I\Sigma_n$. Note that by 2.5 (3), formulas Δ_{n+1} in $B\Sigma_{n+1}$ are closed under bounded quantifiers.

This completes our list of facts. In the sequel we shall present a series of lemmas that proves all the above theorems.

2.8 Lemma. (a) For each φ , $L\varphi \equiv I'\neg\varphi$ is provable in predicate logic (trivial). Thus $I'\Sigma_n \Leftrightarrow L\Pi_n$ and $I'\Pi_n \Leftrightarrow L\Sigma_n$.

(b) $I\Sigma_0 \Leftrightarrow L\Sigma_0 \Leftrightarrow I'\Sigma_0$.

Proof. The second equivalence is obvious from (a); for $I\Sigma_0 \Rightarrow L\Sigma_0$ see 1.22. We prove $L\Sigma_0 \rightarrow I\Sigma_0$. Assume $\varphi \in \Sigma_0$ and let us work in $L\Sigma_0$. Assume

$\varphi(0)$, $(\forall x)(\varphi(x) \rightarrow \varphi(S(x)))$ and $(\exists x)\neg\varphi(x)$. Then there is an x such that $\neg\varphi(x) \& (\forall y < x)\varphi(y)$. But $x \neq 0$ since $\varphi(0)$, thus $x = S(y)$ for some y ; by Q' , $y < x$, thus $\varphi(y)$ and therefore $\varphi(x)$ – a contradiction. $\square$

2.9 Lemma. In $B\Sigma_n$, Σ_n -formulas as well as Π_n -formulas are closed under bounded quantification.

Proof. Evident for $n = 0$. Assume the lemma for n and consider $B\Sigma_{n+1}$ and a Σ_{n+1} formula $(\exists y)\varphi(x, y)$. Then

$$(*) \quad B\Sigma_{n+1} \vdash (\forall x \leq u)(\exists y)\varphi(x, y) \equiv (\exists v)(\forall x \leq u)(\exists y \leq v)\varphi(x, y)$$

and the formula following $(\exists v)$ on the right hand side of $(*)$ is Π_n in $B\Sigma_n$ by the induction hypothesis. $\square$

2.10 Lemma. $B\Sigma_{n+1} \Leftrightarrow B\Pi_n$.

Proof by contraction of quantifiers, cf. 1.45. $\square$

2.11 Lemma. $I\Sigma_{n+1} \Rightarrow B\Sigma_{n+1}$.

Proof. For $n = 0$ see 1.44–1.45; proceed by induction on n . Assume the lemma for $n - 1$, let $\varphi \in \Pi_n$. In $I\Sigma_{n+1}$ assume $(\forall x \leq u)(\exists y)\varphi(x, y)$. Prove

$$(\forall v \leq u)(\exists t)(\forall x \leq v)(\exists y \leq t)\varphi(x, y)$$

by induction on v ; note that, by $B\Sigma_n$, the formula $(\forall x \leq v)(\exists y \leq t)\varphi$ is Π_n in our theory (call it α) and consequently the formula $v \leq u \rightarrow (\exists t)\alpha$ is Σ_{n+1} in $I\Sigma_{n+1}$ so that induction can be used. $\square$

2.12 Lemma. (1) $I\Sigma_n \Leftrightarrow I'\Sigma_n$ and similarly for Π_n and $\Sigma_0(\Sigma_n)$. Thus $I\Sigma_n \Leftrightarrow L\Pi_n$ and $I\Pi_n \Leftrightarrow L\Sigma_n$.

(2) $I\Sigma_n \Leftrightarrow I\Pi_n$.

Proof. (1) For $n = 0$ see above. Consider $n > 0$. First work in $I\Sigma_n$. Let $(\forall x)[(\forall y < x)\varphi(y) \rightarrow \varphi(x)]$; then we get $(\forall x)(\forall y < x)\varphi(y)$ by induction, observing that $(\forall y < x)\varphi(y)$ is Σ_n in $I\Sigma_n$.

Conversely, work in $I'\Sigma_n$; note that we have $I\Sigma_0$, thus basic properties of $<$, S are provable. Assume $\varphi(0)$ and $(\forall x)(\varphi(x) \rightarrow \varphi(S(x)))$. We prove $(\forall x)[(\forall y < x)\varphi(y) \rightarrow \varphi(x)]$ by cases: If $x = 0$ we have $\varphi(x)$. If $x > 0$ then $x = S(z)$ for some z , thus $(\forall y < x)\varphi(y) \rightarrow \varphi(z) \rightarrow \varphi(S(z)) \rightarrow \varphi(x)$. By order induction we get $(\forall x)\varphi(x)$.

(2) Trivial for $n = 0$. Let $\varphi(x)$ be Π_n and work in $I\Sigma_n$; assume $\varphi(0)$, $(\forall x)(\varphi(x) \rightarrow \varphi(x + 1))$ and $\neg\varphi(a)$. Prove the following by induction on z , observing that the formula for which induction is used is Σ_n :

$$(\forall z)(z \leq a \rightarrow \neg\varphi(a - z)).$$

(cf. 1.13–1.16; in greater detail we could write $(\forall z)(z \leq a \rightarrow (\exists u \leq a)(z + u = a \ \& \ \neg\varphi(u))$). But then we get $\neg\varphi(a - a)$, thus $\neg\varphi(0)$, a contradiction). $\square$

2.13 Lemma. Let $n \geq 1$. (a) Each Σ_n formula is piecewise coded in $I\Sigma_n$. (b) Each $\Sigma_0(\Sigma_n)$ formula is piecewise coded in $I\Sigma_n$.

Proof. (a) $I\Sigma_n \vdash (\forall z)(\exists q)(q \text{ } z\text{-piece of } \varphi)$. Indeed, q is the least y such that y is a mapping of $\langle z \rangle^k$ into $\{0, 1\}$ and

$$(\forall x_0 < z), \dots, (\forall x_k < z)(\varphi(x_0 \dots x_k) \rightarrow q(\langle x_0 \dots x_k \rangle) = 1);$$

q exists by $L\Pi_n$.

This proves (a).

(b) Let $\varphi, \psi \in \Sigma_n$; then $I\Sigma_n$ proves the following: If φ, ψ have a z -piece then $\varphi \ \& \ \psi, \neg\varphi, (\exists x \leq y)\varphi$ have a z -piece.

This follows easily by $\Sigma_0^{exp}(exp)$ – comprehension. Assertion (b) follows from the above by induction of the complexity on the $\Sigma_0(\Sigma_n)$ formula in question. $\square$

2.14 Lemma. $I\Sigma_n \Rightarrow I\Sigma_0(\Sigma_n)$.

Proof. Trivial for $n = 0$, thus let $n > 0$, let $\varphi \in \Sigma_0(\Sigma_n)$ and work in $I\Sigma_n$: we show $L\varphi$. Assume $\varphi(a)$ and let q be a z -piece of φ (with respect to the variable in question), $z > a$. (Use 2.13.) By $I\Sigma_1$, let i be the least number such that $(q)_i = 1$; then obviously $(q)_i$ is the least element x such that $\varphi(x)$. We have proved $L\Sigma_0(\Sigma_n)$; the proof of $I\Sigma_0(\Sigma_n)$ from $L\Sigma_0(\Sigma_n)$ is the same as the proof of 2.8 (b). $\square$

2.15 Lemma. $B\Sigma_{n+1} \Rightarrow I\Sigma_n$.

Proof. Trivial for $n = 0$; thus assume $n > 0$ and work in $B\Sigma_{n+1}$. Note that we may assume $I\Sigma_{n-1}$ (induction on n). Let $\varphi(x)$ be Σ_n , $\varphi(x) \equiv (\exists z)\psi(x, z)$. In $B\Sigma_{n+1}$ assume $\varphi(0)$, $(\forall x)(\varphi(x) \rightarrow \varphi(x + 1))$. Let a be given; we prove $(\forall x \leq a)\varphi(x)$. We have the following:

$$\begin{aligned} & (\forall x)(\varphi(x) \rightarrow \varphi(x + 1)) \rightarrow \\ & (\forall x \leq a)[(\exists z)\psi(x, z) \rightarrow (\exists w)\psi(x + 1, w)] \rightarrow \\ & (\forall x \leq a)(\exists w)(\forall z)(\psi(x, z) \rightarrow \psi(x + 1, w)) \rightarrow \\ & (\exists v)(\forall x \leq a)(\exists w \leq v)(\forall z)(\psi(x, z) \rightarrow \psi(x + 1, w)). \end{aligned}$$

Fix such a v ; we may also assume $(\exists w \leq v)\varphi(0, w)$. Then prove $(\forall x \leq a)(\exists w \leq v)\psi(x, w)$ by induction on x using $I\Sigma_{n-1}$. $\square$

2.16 Lemma. $B\Sigma_{n+1} \Rightarrow L\Delta_{n+1}$.

Proof. Work in $B\Sigma_{n+1}$. Assume $(\forall x)[(\exists y)\varphi(x, y) \equiv \neg(\exists z)\psi(x, z)]$, where φ, ψ are Π_n ; furthermore, assume $(\exists y)\varphi(a, y)$. We have

$$\begin{aligned} & (\forall x)(\exists y)(\varphi(x, y) \vee \psi(x, y)), \text{ hence for some } w, \\ & (\forall x \leq a)(\exists y \leq w)(\varphi(x, y) \vee \psi(x, y)) \text{ (by } B\Sigma_{n+1}), \text{ thus} \\ & (\forall x \leq a)((\exists y)\varphi(x, y) \rightarrow (\exists y \leq w)\varphi(x, y)). \end{aligned}$$

Then we have $(\exists y \leq w)\varphi(a, y)$ and we may use $I\Sigma_n$ (preceding lemma), thus $L\Pi_n$ (2.11). Thus there is a least $a' \leq a$ such that $(\exists y \leq w)\varphi(a', y)$; but then a' is the least number such that $(\exists y)\varphi(a', y)$. This completes the proof. $\square$

2.17 Lemma. $L\Delta_{n+1} \Rightarrow B\Sigma_{n+1}$.

Proof. By induction on n . Work in $L\Delta_{n+1}$, let $(\forall x \leq a)(\exists y)\theta(x, y)$, θ being Π_n . Put $\varphi(x) \equiv x \leq a \& (\exists v)[\theta(x, v) \& (\forall u < v)\neg\theta(x, u) \& (\forall z \text{ between } x, a)(\exists y \leq v)\theta(z, y)]$; thus $\varphi(x)$ means that the least witness for x majorizes the least witness for all z such that $x \leq z \leq a$. Observe that φ is Σ_{n+1} ; for $n = 0$ this is evident, and for $n > 0$ it follows by $B\Sigma_n$. But $\varphi(x)$ is equivalent to the following formula $\psi(x)$: $\psi(x) \equiv x \leq a \& (\forall v)[\theta(x, v) \rightarrow (\forall z \text{ between } x, a)(\exists y \leq v)\theta(z, y)]$, thus: each witness for x majorizes the witnesses in question. Now $\psi(x)$ is Π_{n+1} (for analogous reasons) and therefore $L\Delta_{n+1}$ applies: let x_0 be the least element satisfying φ and $\theta(x_0, y_0)$.

Claim. $(\forall x \leq a)(\exists y \leq y_0)\theta(x, y)$. Indeed, the statement holds for x between x_0 and a by definition; if there is an $x < x_0$ such that $(\forall y \leq y_0)\neg\theta(x, y)$ then take the largest such x possible — x is the least element such that

$$x < x_0 \& (\forall x' < x_0)(x < x' \rightarrow (\exists y < y_0)\theta(x, y)),$$

which is Δ_{n+1} . But then we have $\varphi(x') \& x' < x$ — a contradiction. $\square$

2.18 Remark. (1) The proof of $L\Delta_{n+1} \Rightarrow I\Delta_{n+1}$ is easy and left to the reader (cf. 2.8).

(2) The reader may check that Theorem 2.4 follows from 2.8, 2.12 and 2.14; furthermore, Theorem 2.5 follows from 2.9, 2.10, 2.11, 2.15, 2.16, 2.17 and 2.18 (1). In addition, Theorem 2.7 (1) is proved in 2.13. Thus it remains to prove 2.7 (2); this is done in the following lemma.

2.19 Lemma. If $n \geq 1$ and φ is Δ_{n+1} in $B\Sigma_{n+1}$ then φ is piecewise coded in $B\Sigma_{n+1}$.

Proof. This is proved similarly to 2.13 (a): given z , the desired z -piece of $\varphi(x)$ is the least q such that q is a finite mapping, $\text{dom}(q) = (< z)^k$, $\text{range}(q) \subseteq \{0, 1\}$ and $(\forall x_0 < z) \dots (\forall x_k < z)(\varphi(x_0 \dots x_k) \rightarrow q(\langle x_0 \dots x_k \rangle) = 1)$. The condition in question is easily shown to be Δ_{n+1} in $B\Sigma_{n+1}$, thus $L\Delta_{n+1}$ applies. This completes the proofs of our theorems. $\square$

(b) Further Principles and Facts About Fragments

2.20 Principles – continued. Let us introduce three new principles:

Strong collection $S\varphi$:

$$(\forall u)(\exists v)(\forall x \leq u)[(\exists y)\varphi(x, y) \rightarrow (\exists y \leq v)\varphi(x, y)]$$

Regularity $R\varphi$:

$$(Cx)(\exists y \leq u)\varphi(x, y) \rightarrow (\exists y \leq u)(Cx)\varphi(x, y)$$

where (Cx) is $(\forall w)(\exists x > w)$ – the quantifier “there are unboundedly many”

Pigeon hole principle $PHP(\varphi)$:

$$(\forall u)\neg[(\forall x \leq u)(\exists! y \leq u+1)\varphi(x, y) \ \& \ (\forall y \leq u+1)(\exists! x \leq u)\varphi(x, y)]$$

We offer the reader the following verbal reformulations: $S\varphi$ may be understood as saying that the partial multivalued function on $(< u)$ defined by φ is not cofinal on its domain. $R\varphi$ says that if unboundedly many x 's have a value beneath u then there is a $y < u$ which is a value of unboundedly many x 's. $PHP(\varphi)$ just says that φ cannot define a one-one mapping of $(< u)$ onto $(< u+1)$.

2.21 Principles – completed. We shall introduce three more principles, all asserting the existence of some sequences. Since we defined sequences in $I\Sigma_0(exp)$ we are obliged to relate these principles to theories containing $I\Sigma_0(exp)$; our choice will be $I\Sigma_1$.

The first principle is called the *finite axiom of choice* ($FAC(\varphi)$) and claims that if a multivalued function is total beneath u then there is a sequence s which selects for each $x < u$ one of the values of x (i.e. $(s)_x$ is the selected value).

$$(\forall x \leq u)(\exists y)\varphi(x, y) \rightarrow (\exists s)(Seq(s) \ \& \ lh(s) = u+1 \\ \& \ (\forall x \leq u)\varphi(x, (s)_x).$$

The last two principles concern approximations of functions. Let us begin with informal formulations; we shall show how to obtain the corresponding formulas. Say that φ defines a partial function if $(\forall x, y, z)(\varphi(x, y) \ \& \ \varphi(x, z) \rightarrow y = z)$ (abbreviate this $PFUN(\varphi)$); say that φ defines a total function if $(\forall x)(\exists! y)\varphi(x, y)$ (abbreviate this $TFUN(\varphi)$). Write $y = F(x)$ for $\varphi(x, y)$ for a moment. Call a finite sequence s an *approximation* of F if for each i such that $i+1 < lh(s)$ and each $x \leq (s)_i$, $y = F(x)$ implies $y \leq (s)_{i+1}$. The principle of approximation for a *total* function says that if φ defines a total function then this function has arbitrarily long approximations; similarly for *partial* functions. Write $Approx_\varphi(s)$ for

$$Seq(s) \ \& \ (\forall i)(\forall x)(i+1 < lh(s) \ \& \ x \leq (s)_i \ \& \ \varphi(x, y) \rightarrow y \leq (s)_{i+1}).$$

The principles are as follows:

$$\begin{aligned}(P_\varphi) \quad & PFUN(\varphi) \rightarrow (\forall z)(\exists s) Approx_\varphi(s) \& lh(s) = z), \\(T_\varphi) \quad & TFUN(\varphi) \rightarrow (\forall z)(\exists s) Approx_\varphi(s) \& lh(s) = z).\end{aligned}$$

Note that both (P_φ) and (T_φ) are meaningful in $I\Sigma_1$ (even in $I\Sigma_0(exp)$; *Seq* is used). This completes our list of principles; obviously, for each φ all the above principles are true in N .

2.22 Theories. (1) Theories based on S , R and PHP are assumed to contain $I\Sigma_0$:

$$\begin{aligned}S\Gamma &= I\Sigma_0 \cup \{S\varphi \mid \varphi \in \Gamma\}, \\R\Gamma &= I\Sigma_0 \cup \{R\varphi \mid \varphi \in \Gamma\}, \\PHP(\Gamma) &= I\Sigma_0 \cup \{PHP(\varphi) \mid \varphi \in \Gamma\}.\end{aligned}$$

Here Γ stands for Σ_n , Π_n and possibly $\Sigma_0(\Sigma_n)$.

(2) Theories based on FAC , T , P are assumed to contain $I\Sigma_1$:

$$\begin{aligned}FAC(\Gamma) &= I\Sigma_1 \cup \{FAC(\varphi) \mid \varphi \in \Gamma\} \\T(\Gamma) &= I\Sigma_1 \cup \{T\varphi \mid \varphi \in \Gamma\} \\P(\Gamma) &= I\Sigma_1 \cup \{P\varphi \mid \varphi \in \Gamma\}.\end{aligned}$$

Thus we may use the notion of finite sequences. Note that in Chap. V a theory of finite sequences in $I\Sigma_0$ will be elaborated; having this we could discuss also the principles FAC , T , P over $I\Sigma_0$. But we shall not investigate this now. Our results are contained in the following three theorems:

2.23 Theorem. For each natural n :

- (1) $S\Sigma_{n+1} \Leftrightarrow S\Pi_n \Leftrightarrow I\Sigma_{n+1}$
- (2) $PHP(\Sigma_0(\Sigma_n)) \Leftrightarrow I\Sigma_{n+1}$
- (3) for $n \geq 1$, $PHP(\Sigma_{n+1}) \Leftrightarrow B\Sigma_{n+1}$; furthermore, $PHP(\Sigma_1) \Rightarrow B\Sigma_1$.
- (4) $R\Sigma_{n+1} \Leftrightarrow R\Pi_n \Leftrightarrow B\Pi_{n+1} \Leftrightarrow B\Sigma_{n+2}$.

2.24 Theorem.

- (1) For $n \geq 1$, $B\Sigma_{n+1} \Leftrightarrow FAC(\Sigma_{n+1})$; and $I\Sigma_1 \Leftrightarrow FAC(\Sigma_1)$.
- (2) For $n \geq 1$, $T\Sigma_{n+1} \Rightarrow P\Sigma_{n+1} \Rightarrow T\Sigma_n$.
- (3) For $n \geq 0$, $I\Sigma_{n+1} \Leftrightarrow T\Sigma_{n+1} \Leftrightarrow T\Pi_n$, and $P\Sigma_{n+1} \Leftrightarrow P\Pi_n$.

2.25 Theorem. Each $\Sigma_0(\Sigma_n)$ formula is Δ_{n+1} in $I\Sigma_n$.

2.26 Remark. The structure of our fragments is visualized in Fig. 1. We shall show in Chap. IV that all inclusions are strict and that theories incomparable in the figure are incomparable as theories, i.e. neither includes the other.

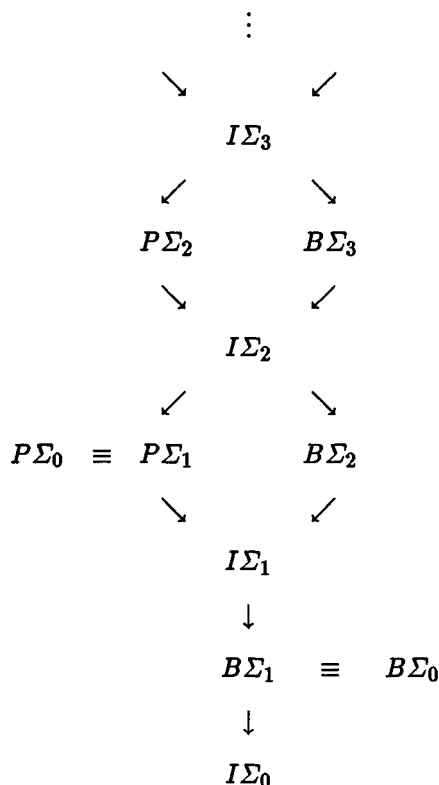


Fig. 1.

The subsequent series of lemmas proves our three theorems.

Overview. 2.23 (1) proved in 2.27–2.29, (2–3) in 2.30–2.33, (4) in 2.36–2.44 and uses 2.24 (1); 2.24 (1) proved in 2.35, (2–3) in 2.45–2.48 and uses 2.37; 2.25 is proved in 2.49–2.50 and uses 2.23 (1).

2.27 Lemma. $I\Sigma_{n+1} \Rightarrow S\Pi_n$.

Proof. Let φ be Π_n and work in $I\Sigma_{n+1}$. Let u be given; let q be the $(u+1)$ -piece of $(\exists y)\varphi(x, y)$ (cf. 2.13). We prove $(\exists v)(\forall x \leq a)((q)_x = 1 \rightarrow (\exists y \leq v)\varphi(x, y))$. This is achieved easily by proving the following sentence by induction on z :

$$(*) \quad z \leq u \rightarrow (\exists v)(\forall x \leq z)((q)_x = 1 \rightarrow (\exists y \leq v)\varphi(x, y)).$$

This is clear for $z = 0$. Let v satisfy $(*)$ w.r.t. z and let $z+1 \leq u$; if $\neg(\exists y)\varphi(z+1, y)$ then v satisfies $(*)$ for $u+1$; if $\varphi(z+1, y)$, then $(*)$ holds for $z+1$ and $\max(v, y)$ instead of z, v . This completes the proof. $\square$

2.28 Lemma. $S\Pi_n \Rightarrow S\Sigma_{n+1}$.

Proof. Contracting quantifiers (cf. 2.10). □

2.29 Lemma. $S\Pi_n \Rightarrow I\Sigma_{n+1}$.

Proof. By induction on n . Recall that $S\Pi_n$ contains $I\Sigma_0$. Let φ be Π_n . Work in $S\Pi_n$; assume $(\exists y)\varphi(0, y)$ and $(\forall x)((\exists y)\varphi(x, y) \rightarrow (\exists y)\varphi(x+1, y))$. Take any u and use $S\Pi_n$ to get a v such that $(\forall x \leq u)((\exists y)\varphi(x, y) \rightarrow (\exists y \leq v)\varphi(x, y))$. Then use $I\Sigma_n$ to prove $(\forall x \leq u)(\exists y \leq v)\varphi(x, y)$. □

2.30 Lemma. $I\Sigma_{n+1} \Leftrightarrow PHP(\Sigma_0(\Sigma_{n+1}))$.

Proof. $(\Rightarrow)$ follows by Lemma 2.13 (b): if φ defines a mapping of $u+1$ into u then its $(u+1)$ -piece is a finite set function and using comprehension we obtain the mapping in question as a finite set. But $I\Sigma_1$ proves the pigeon hole principle for finite sets, see 1.41.

Conversely, assume $PHP(\Sigma_0(\Sigma_{n+1}))$ and let φ be a Σ_{n+1} -formula violating $I\Sigma_{n+1}$: let a be such that

$$\varphi(0) \& (\forall x)(\varphi(x) \rightarrow \varphi(x+1) \& \neg\varphi(a)).$$

Define, for $x \leq a$, $f(x) = x-1$ if $\neg\varphi(x)$ and $f(x) = x$ otherwise. Then f is $\Sigma_0(\Sigma_{n+1})$ -definable and violates $PHP(\Sigma_0(\Sigma_{n+1}))$. □

2.31 Lemma. $PHP(\Sigma_{n+1}) \Rightarrow B\Sigma_{n+1}$.

Proof. We shall construct the proofs by induction on n ; this allows us to assume $B\Sigma_n$ if $n > 0$. Let n be given and work in $PHP(\Sigma_{n+1})$.

(1) First prove $I\Sigma_n$. This is immediate if $n = 0$ (since $PHP(\Sigma_{n+1})$ extends $I\Sigma_0$). If $n > 0$ imitate the proof of 2.30 $(\Leftarrow)$, φ being now Σ_n ; by $B\Sigma_n$, the function f is now Σ_{n+1} definable and $PHP(\Sigma_{n+1})$ gives the result.

(2) Now assume that $B\varphi$ is violated for a $\varphi \in \Pi_n$; we want to deduce a contradiction, which will complete the proof. Put $\varphi'(x, y) \equiv (\exists z \leq y)\varphi(x, z) \& y = (x, z)$; then $B\varphi'$ fails and $x_1 < x_2 \leq u \& \varphi(x_1, y_2)$ implies $y_1 = y_2$. Thus $(\forall x \leq u)(\exists y)\varphi'(x, y)$ but $(\forall v)(\exists x \leq v)(\forall y \leq v)\neg\varphi'(x, y)$. By $I\Sigma_n$, for each $x \leq u$ there is a least y such that $\varphi'(x, y)$; denote it by $m(x)$. Clearly, m is one-one on $(\leq u)$. For $x, x' \leq u$ put $x' = F(x)$ if $m(x) < m(x')$ and $\neg(\exists x'' < u)(m(x) < m(x'') < m(x'))$.

(3) We show that F is Σ_{n+1} . Indeed, since m is $\Sigma_0(\Sigma_n)$ and we have $I\Sigma_n$, m is piecewise coded; thus for each z , $\{(x, y) \mid x \leq u \& y \leq z \& y = m(x)\}$ is a finite set. (Say that the set codes m on $(\leq u) \times (\leq z)$.) Thus $x' = F(x)$ is equivalent to $(\exists q)(q \text{ codes } m \text{ on } (\leq u) \times (\leq z) \& (\exists y, y' \leq z)[q(x) = y \& q(x') = y' \& \neg(\exists x' \leq u)(\exists y'' \leq z)(q(x'') = y'' \& y < y'' < y')])$.

The formula " q codes m on $(\leq u) \times z$ " is Δ_{n+1} , the second conjunct is Δ_1 ; thus the whole is Σ_{n+1} .

(4) F is total on $(\leq u)$ and there is an $x_0 \leq u$ such that F maps $(\leq u)$ onto $(\leq u) - \{x_0\}$. Indeed, there is the least y such that $(\exists x \leq u)\varphi'(x, y)$ (by $I\Sigma_n$; the formula is Π_n in $I\Sigma_n$). If $\varphi(x_0, y)$ then $y = m(x_0)$ and $y \notin \text{range}(F)$. For each $x' \leq u$ different from x_0 there is an $x \leq u$ such that $x' = F(x)$: to see this assume $x' \neq x_0$, $y = m(x')$ and let q code m on $(\leq u) \times (\leq y)$. The desired x is easily obtained from q using $I\Sigma_1$.

(5) F is one-one: trivial from the definition. Thus F maps $(\leq u)$ one-one onto $(\leq u) - \{x_0\}$ and F is Σ_{n+1} . Changing F inessentially we may assume $x_0 = u$, thus $\text{range}(F) = (\leq u - 1)$ and we have a contradiction with $PHP(\Sigma_{n+1})$. $\square$

2.32 Lemma. $(B\Sigma_{n+1} + PHP(\Sigma_n)) \Rightarrow PHP(\Sigma_{n+1})$.

Proof. If $\theta(x, y) \equiv (\exists z)\psi(x, y, z)$ is Σ_{n+1} and defines in $B\Sigma_{n+1}$ a 1-1-mapping of $(\leq a)$ to $(\leq a - 1)$ then find a t such that the formula $(\exists z < t)\psi(x, y, z)$ (which is Π_n in $B\Sigma_{n+1}$) defines the same mapping. $\square$

2.33 Lemma. For $n \geq 1$, $B\Sigma_{n+1} \Rightarrow PHP(\Sigma_{n+1})$

Proof. In the theory in question, we have $I\Sigma_n$, which implies $PHP(\Pi_n)$ by 2.30, and 2.32 gives the result. $\square$

2.34 Remark. Note that 2.23 (1)–(3) has been proved. Before we prove 2.23 (4) let us discuss FAC .

2.35 Lemma. $I\Sigma_1 + B\Pi_m \Leftrightarrow FAC(\Pi_m)$.

Remark. If $m = 0$ then $I\Sigma_1 + B\Pi_m$ is $I\Sigma_1$; if $m > 0$ it is $B\Sigma_m$.

Proof. $(\Rightarrow)$ In $I\Sigma_1 + B\Pi_m$ assume $(\forall x \leq u)(\exists y)\varphi(x, y)$, $\varphi \in \Pi_m$; thus for some v we have $(\forall x \leq u)(\exists y \leq v)\varphi(x, y)$. Prove the following formula by induction on z :

$$(\forall z \leq u)(\exists s \leq 2^{(u+v+1)^2})(lh(s) = x + 1 \ \& \ (\forall i \leq x)\varphi(i, (s)_i))$$

(cf. the remark in 1.43). Observe that the formula in question is Δ_1 for $m = 0$ and is Δ_m for $m > 0$; thus the corresponding induction axiom is at our disposal.

$\Leftarrow$ is evident. $\square$

Remark. Observe that $FAC(\Pi_m) \Rightarrow FAC(\Sigma_{m+1})$ (cf. 2.10); thus 1.24 (1) is proved.

2.36 Definition. Given a formula φ , the following formula (that can be denoted $MFUNS(\varphi)$) says that φ defines a monotone function on an initial segment:

$$\begin{aligned} & (\forall x, y_1, y_2)(\varphi(x, y_1) \& \varphi(x_1, y_2) \rightarrow y_1 = y_2) \\ & (\forall x_1, y_1, x_2, y_2)(x_1 < x_2 \& \varphi(x_1, y_1) \& \varphi(x_2, y_2) \rightarrow y_1 < y_2) \\ & (\forall x_1, x_2, y_2)(x_1 < x_2 \& \varphi(x_2, y_2) \rightarrow (\exists y_1)\varphi(x_1, y_1)). \end{aligned}$$

2.37 Lemma. For each formula $\theta(x, y) \in \Pi_m$ there is a $\psi_0 \in \Pi_m$ such that $(I\Sigma_1 + B\Pi_m)$ proves the following:

(a) $(\forall x < a)(\exists y)\theta \equiv (\forall x < a)(\exists y)\psi_0$.

(b) ψ_0 defines an increasing function on an initial segment. (Note once more that if $m = 0$ then the theory in question is $I\Sigma_1$; if $m > 0$ then it is $B\Pi_m$.)

Proof. For $m = 0$, ψ_0 says: y is the function associating with each $x_0 \leq x$ the least y_0 such that $\theta(x_0, y_0)$. For $m > 0$ let $\theta(x_0, y_0)$ be $(\forall z)\lambda(x_0, y_0, z)$.

We make the following definitions. A *superwitness* for x is a pair (y, s) where $\theta(x, y)$ (i.e. y witnesses $(\exists y)\varphi(x, y)$) and s is a sequence of length y such that for each $z < y$, $(s)_z$ is the smallest u such that $\neg\lambda(x, z, u)$ (this sequence witnesses for each $z < y$ that $\neg\theta(x, z)$ holds; altogether, (y, s) witnesses that y is the least element satisfying $\theta(x, y)$). Write $SW(x, (y, s))$ for “ y is the superwitness for x ” and observe $SW(x, (y, s)) \in \Pi_m$. Now define $SSW(x, q)$ (q is a super-superwitness for x) if q is a sequence of length $x + 1$ such that for each $z \leq x$, $(q)_z$ is the superwitness for z . Observe that $B\Pi_m$ proves the following

- (1) $SW(x, t) \& SW(x, t') \rightarrow t = t'$;
- (2) $SSW(x, t) \& SSW(x, t') \rightarrow t = t'$;
- (3) $x < x' \& SSW(x', t') \rightarrow (SSW(x, t) \equiv t \text{ is the restriction of } t' \text{ to } (\leq x))$;
- (4) $(\forall x \leq a)(\exists y)\theta(x, y) \equiv (\exists t)SSW(a, t)$.

To prove (4), observe that $\Leftarrow$ is trivial; let us prove $\Rightarrow$. Assume $(\forall x \leq a)(\exists y)\theta(x, y)$; then $(\forall x \leq a)(\exists w)SW(x, w)$, since (i) by $L\Pi_m$, there exists the least y such that $\theta(x, y)$, (ii) again by $L\Pi_m$, for each $z < y$ exists the least u such that $\neg\lambda(x, z, u)$ and by $FAC(\Pi_m)$, there is a sequence s associating with each $z < y$ this least u . Applying $FAC(\Pi_m)$ to $(\forall x \leq a)(\exists w)SW(x, w)$ we gain directly the sequence which is the desired super-superwitness. The lemma now follows easily. $\square$

2.38 Definition ($I\Sigma_1$). If $\psi_0(x, y)$ defines an increasing function F on an initial segment then call a sequence s a *code of the primitive recursive iteration of F* if $(s)_0 = 0$ and for each x such that $x + 1 < lh(s)$ we have $(s)_{x+1} = F((s)_x)$. (In symbols: $CPRI_{\psi_0}(s)$.)

2.39 Lemma. Let ψ_0 be Π_m ; ($I\Sigma_1 + B\Pi_m$) proves the following: If ψ_0 defines an increasing function F on an initial segment then

(a) $(\forall x < a)(\exists y)\psi_0 \rightarrow (\exists t > a)CPRI_{\psi_0}(t)$

(b) Let $K(a)$ be the constant sequence of length a whose each member is a . Then

$$(\exists t > K(a))CPRI_{\psi_0}(t) \rightarrow (\forall x < a)(\exists y)\psi_0.$$

Proof. (a) Note that $CPRI_{\psi_0}(t)$ is Π_m and we have $L\Pi_m(L\Pi_1$ for $m = 0$.) Let s be the maximal code of the primitive recursive iteration of F such that $s \leq a$; let $u = lh(s) - 1$ and $v = (s)_u$. Clearly, $v < a$ and if w is such that $\psi_0(v, w)$ then let t be the concatenation of a with the element w , i.e. $t = s \smallfrown \langle w \rangle$. Then $t > s$, thus $t > a$.

(b) Assume $\neg(\forall x < a)(\exists y)\psi_0$. Then evidently each primitive recursive iteration t of F is less than $K(a)$. $\square$

2.40 Lemma. Let ψ_0 be Π_m ; ($I\Sigma_1 + B\Pi_m$) proves the following: if ψ_0 defines an increasing function on an initial segment then

$$(\forall x)(\exists y)\psi_0 \equiv (Ct)CPRI_{\psi_0}(t).$$

Immediate from the preceding.

2.41 Lemma. For each $\theta(x, y) \in \Pi_m$ there is a $\psi(t) \in \Pi_m$ such that ($I\Sigma_1 + B\Pi_m$) $\vdash (\forall x)(\exists y)\theta \equiv (Ct)\psi$.

Immediate from the preceding.

2.42 Lemma. $B\Sigma_{m+2} \Rightarrow R\Sigma_{m+1}$

Proof. Assume $\neg(\exists y < a)(Cx)\theta(\theta \in \Sigma_{m+1})$. Using $B\Pi_{m+1}$ show $(\exists t)(\forall y < a)(\forall x > t)\neg\theta$, i.e. $\neg(Cx)(\exists y < a)\theta$. $\square$

2.43 Lemma. $R\Pi_0 \Rightarrow I\Sigma_1$.

Proof. Recall that by definition, $R\Pi_0$ contains $I\Sigma_0$. In $R\Pi_0$, assume $(\exists z)\varphi(a, z)$ where $\varphi \in \Sigma_0$. Trivially,

$$(\forall x \leq a)(\exists w)(\forall u > w)[(\exists z < u)\varphi(x, z) \equiv (\exists z < u + 1)\varphi(x, z)].$$

Applying $R\Pi_0$ we get

$$\neg(Cu)(\exists x < a)[(\exists z < u + 1)\varphi(x, z) \& \neg(\exists z < u)\varphi(x, z)]$$

thus

$$(\exists t)(\forall u \geq t)(\forall x < z)[(\exists z < u)\varphi(x, z) \equiv (\exists z < u + 1)\psi(x, z)].$$

Hence

$$(\forall x \leq a)[(\exists z)\varphi(x, z) \equiv (\exists z < t)\psi(x, z)].$$

Thus $I\Sigma_0$ gives the desired minimum. $\square$

2.44 Lemma. $R\Pi_m \Rightarrow B\Pi_{m+1}$

Proof by induction on m ; thus we may assume $B\Pi_m$. (For $m = 0$ this follows from 2.43.) Let $\theta(x, y, z)$ be Σ_m and consider $(\forall z)\theta(x, y, z)$. Work in $(R\Pi_m + B\Pi_m)$; assume

$$(\forall x < a)(\exists y)(\forall z)\theta(x, y, z).$$

By 2.37, let $\psi_0(x, y, z) \in \Pi_m$ be such that ψ_0 defines an increasing function on an initial segment and

$$(\forall q)(\forall y < q)(\exists z)\neg\theta(x, y, z) \equiv (\forall y < q)(\exists z)\psi_0(x, y, z). \quad (*)$$

Then we have the following:

$$\begin{aligned} &(\forall x < a)(\exists y)(\forall z)\theta(x, y, z) \rightarrow && \text{by } (*) \\ &(\forall x < a)\neg(\forall y)(\exists z)\psi_0(x, y, z) \rightarrow && \text{(by 2.40)} \\ &(\forall x < a)\neg(Ct)CPRI_{\psi_0}(x, t) \rightarrow && \text{(by } R\Pi_m) \\ &\neg(Ct)(\exists x < a)CPRI_{\psi_0}(x, t) \rightarrow && \text{(definition of } C) \\ &(\exists q)(\forall t > q)(\forall x < a)\neg CPRI_{\psi_0}(x, t) \rightarrow && \text{(logic)} \\ &(\exists q)(\forall x < a)(\forall t > q)\neg CPRI_{\psi_0}(x, t) \rightarrow && \text{(cf. 2.39)} \\ &(\exists q)(\forall x < a)(\exists y < q)(\forall z)\neg\psi_0(x, y, z) \rightarrow && \text{(by } *) \\ &(\exists q)(\forall x < a)(\exists y < q)(\forall z)\theta(x, y, z). \end{aligned}$$

This completes the proof of $B\Pi_{m+1}$. $\square$

Remark. Note that the proof of 2.23 (4) is complete. We turn to 2.24 (2)–(3).

2.45 Lemma. $I\Sigma_{n+1} \Rightarrow T\Sigma_{n+1}$

Proof. Let φ be Σ_{n+1} and work in $I\Sigma_{n+1} + TFUN(\varphi)$, i.e. assume that φ defines a total function F . Then φ is Δ_{n+1} in our theory. Define

$$G(x) = (\min y > x)(\forall u \leq x)(F(u) \leq y).$$

G is total, monotone and Δ_{n+1} ; show by the usual technique that its primitive recursive iteration $H(0) = 0$, $H(x+1) = G(H(x))$ is total and Δ_{n+1} . Moreover, H is piecewise coded; from this it is easy to conclude that for each x , the restriction of H to $(\leq x)$ is a finite set h ; but h is an approximation of F of length $x+1$. $\square$

2.46 Lemma. $T\Sigma_{n+1} \Rightarrow I\Sigma_{n+1}$.

Proof. Trivial for $n = 0$ by our definition of $T\Sigma_1$; thus assume $n > 0$ and work in $T\Sigma_{n+1} + I\Sigma_n$. First prove $B\Pi_n$. Let ψ be Π_n and assume $(\forall x \leq u)(\exists y)\psi(x, y)$. Define $F(x) = (\min y)\psi(x, y)$ for $x \leq u$, $F(x) = 0$ otherwise; F is Δ_{n+1} in our theory. Let s be an approximation of F , $lh(s) > u + 1$. Then $(s)_u \geq u$, thus $(\forall x \leq u)(\exists y \leq (s)_{u+1})\psi(x, y)$. $B\Pi_n$ follows.

To prove $I\Sigma_{n+1}$ take $\varphi \in \Sigma_{n+1}$, $\varphi(x) \equiv (\exists y)\chi(x, y)$ and assume $\varphi(0)$ and $(\forall x)(\varphi(x) \rightarrow \varphi(x + 1))$. For each x , let $G(x) = y$ if y is a sequence of length $(x + 1)$ and for each $i \leq x$, $(y)_i$ is the least z such that $\chi(x, z)$. G is a function (possibly partial) and is Δ_{n+1} in our theory. If G is bounded, i.e. $y = G(x)$ implies $y \leq w$, then we get $(\forall x)(\exists y \leq w)(y = G(x))$ by $I\Delta_{n+1}$, which is at our disposal (thanks to $B\Sigma_{n+1}$); thus assume G unbounded. Thus $(\forall v)(\exists x, y)(y = G(x) \& (x, y) > v)$. Define $H(v)$ to be the least pair (x, y) such that $y = G(x)$; then H is Δ_{n+1} in our theory and H is total. Let s be an approximation of H of length $z + 1$; then s is a sequence of increasing pairs (x, y) such that $y = G(x)$. Thus if $(x, y) = (s)_z$ then $x \geq z$, $y = G(x)$ and therefore y gives witnesses for each $i \leq x$; thus $(\forall i \leq x)(\exists y)\chi(i, y)$ and $\varphi(x)$ follows. $\square$

2.47 Lemma. $T(\Pi_n) \Leftrightarrow T(\Sigma_{n+1})$ and $P(\Pi_n) \Leftrightarrow P(\Sigma_{n+1})$.

Proof. Assume $T(\Pi_n)$; note that we may also assume $I\Sigma_n$ (for $n = 0$ by definition, for $n > 0$ as the induction hypothesis on n). Let $(\exists z)\theta(x, y, z)$ be Σ_{n+1} and assume that this formula defines a total function F . Define $G(x) = (y, q)$ as: q is a superwitness (for $n = 0$: a witness) of $y = F(x)$ (cf. 2.37), i.e., for some z , s , $q = (z, s)$, we have $\theta(x, y, z)$ and s is a sequence witnessing minimality of z . Then $G(x)$ is Π_n , total and majorizes F ; each approximation of G is an approximation of F . Thus $T(\Pi_n)$ gives $T(\Sigma_{n+1})$.

Now assume $P(\Pi_n)$; then $T(\Pi_n)$, thus $T(\Sigma_{n+1})$. Let F be Σ_{n+1} and partial; as above show that there is a Π_n function G with the same domain majorizing F . Each approximation of G is an approximation of F . $\square$

2.48 Lemma. For $n \geq 1$, $I\Sigma_{n+1} \Rightarrow P\Sigma_n \Rightarrow I\Sigma_n$.

Proof. By 2.47, $P(\Sigma_n) \Leftrightarrow P(\Pi_{n-1})$; so assume $T\Sigma_{n+1}$ and let φ be Π_{n-1} and define a partial function G . Then the trivial totalization of G ($H(x) = G(x)$ if defined, $= 0$ otherwise) is Δ_{n+1} , and thus has approximations of arbitrary length; each of them is an approximation of G .

The second statement is trivial since $P\Sigma_n \Rightarrow T\Sigma_n$. This completes the proof of 2.24 (2)–(3). $\square$

Our last task in this subsection is to prove Theorem 2.25. This is done in the following two lemmas.

2.49 Lemma. For each $\Sigma_0(\varphi)$ -formula Ψ there is a formula Ψ' of the form

$$(*) \quad (Q_1 y_1 \leq z_1) \dots (Q_k y_k \leq z_k) \Psi_0(x_1, \dots, y_1, \dots, z_1, \dots)$$

where $Q_1, \dots, Q_k$ are quantifiers, Ψ_0 is a boolean combination of instances of φ and atomic formulas, the sets of variables $\{x_1, \dots\}$, $\{y_1, \dots\}$, $\{z_1, \dots\}$ are pairwise disjoint and the equivalence $\Psi \equiv \Psi'$ is provable in I_{open} . (We call $(*)$ the *bounded prenex normal form* of Ψ .)

Proof. First show that Ψ is equivalent to a formula Ψ_1 consisting of a block of bounded quantifiers followed by a boolean combination of instances of φ and atomic formulas. To this end check that analogues of the following hold for bounded quantifiers: renaming of bound variables, de Morgan rules and factoring out a quantifier Q from the antecedent (succedent) of an implication whose succedent (antecedent) does not contain free occurrence of the variable bound by Q . *Caution:* it is important that bounded quantifiers use non-strict inequalities; i.e. if u is not free in χ then

$$(\exists u \leq v)(\chi \rightarrow \varphi) \equiv (\chi \rightarrow (\exists u \leq v)\varphi) \quad \text{is provable in } I_{\text{open}} \text{ but}$$

$$(\exists u < v)(\chi \rightarrow \varphi) \equiv (\chi \rightarrow (\exists u < v)\varphi) \quad \text{is not}$$

(think of $v = 0$; we need $(\forall v)(\exists u)(u \leq v)$ to be provable).

Ψ_1 has a *clash of variables* if there are variables x, y, z such that Ψ_1 has the form

$$\dots (Q_1 z \leq y) \dots (Q_2 x \leq z) \dots \Psi_0,$$

i.e. z is both the bound in one quantifier and the quantified variable in a quantifier further out. It remains to show that Ψ_1 can be made clash-free. To this end let Ψ_2 be a subformula of Ψ_1 having a clash and such that each proper subformula of Ψ_2 is clash-free. Thus Ψ_2 is

$$(*) \quad (Q y_1 \leq z_1) \dots (\forall y_i \leq y_1) \dots \Psi$$

and similarly for $\exists$ instead of $\forall$. But $(*)$ is equivalent in I_{open} to

$$(Q y_1 \leq z) \dots (\forall y_i \leq z) \dots (y_i \leq y_1 \rightarrow \Psi_0)$$

and similarly for $\exists$.

Iterated use of this procedure gives the result. □

2.50 Lemma. If $\varphi \in \Sigma_n$ and Ψ is $\Sigma_0(\varphi)$ then Ψ is Δ_{n+1} in $I\Sigma_n$.

Proof. Trivial for $n = 0$; thus assume $n > 0$. Assume ψ has the form guaranteed by 2.49 and apply propositional calculus to Ψ_0 ; then we get an equivalent formula Ψ'' of the form

$$(Q y \leq z) (\wedge (\sigma_i \vee \pi_i))$$

where σ_i are Σ_n and π_i are Π_n . Write σ_i as $(\exists y)\sigma_{0i}(x, y)$.

Using $S\Sigma_n$, bound all $(\exists y)$'s:

$$(\exists q)(\forall x, \dots \leq \max(\mathbf{z})) \wedge (\sigma_i \rightarrow (\exists y < q)\sigma_{0i}(\mathbf{x}, y)).$$

Then Ψ'' is equivalent in I_{open} to

$$(\exists q)(Q_1, x_1 \leq z_1) \dots (Q_k x_k \leq z_k) (\wedge (\exists y \leq q)\sigma_{0i} \vee \pi_i),$$

the conjunction is Π_n in $I\Sigma_n$ and the whole last formula is Σ_{n+1} in $I\Sigma_n$. But this suffices since $\neg\Psi$ is also $\Sigma_0(\Sigma_n)$ and therefore Σ_{n+1} in $I\Sigma_n$. This completes the proof of the lemma (and of 2.25). $\square$

(c) Finite Axiomatizability; Partial Truth Definitions for Relativized Arithmetical Formulas

The aim of the present subsection is two-fold: first, to show that for $n > 0$, $I\Sigma_n$, $B\Sigma_{n+1}$ and $P\Sigma_n$ are finitely axiomatizable; we show this using partial truth definitions elaborated in Sect. 1 (d). Second, we shall extend our possibility of dealing with infinite sets inside fragments of arithmetic by showing that, under some conditions, we may speak in $I\Sigma_1$ of sets Σ_n -defined from a given set (and quantify over such sets). This will give us very useful means of expression.

2.51 Discussion. Let us survey our sets and membership in arithmetic. First, we defined in $I\Sigma_0(\exp)$ the membership predicate $\in$ with respect to which numbers behave like hereditarily finite sets; in particular, $I\Sigma_0(\exp)$ proves comprehension for $\Sigma_0^{\exp}(\exp)$ formulas. Furthermore, given any formula $\varphi(x)$ with just one variable, we may introduce a constant (A , say) for the set of all numbers satisfying φ together with an *ad hoc* membership predicate $\in$ such that $x \in A$ just means $\varphi(x)$. Analogously for formulas with more variables – we may introduce a constant for the *relation* defined by φ . In particular, assuming that $\varphi(x, y)$ defines a total function (which we denoted $TFUN(\varphi)$) we may write $y = F(x)$ for $(x, y) \in F$, which in turn just means $\varphi(x, y)$. More generally, our φ may contain free variables distinct from those displayed; they act as parameters. In this case we may consider the parameters to be just *codes* for the corresponding classes; if φ is $\varphi(x, par)$ then $x \in X_{par}$ (or even $x \in par$) means simply $\varphi(x, par)$. A particular case of this is given by partial truth definitions: if $Sat(z, e)$ is a formula which is a partial satisfaction for Γ -formulas (e.g. Σ_n -formulas) then we may take Γ -formulas with exactly one free variable for codes of Γ -sets as we did in Sect. 1 (d). In this last case we have double profit: *first*, we may quantify over Γ -sets and *second*, we have “it’s snowing”-it’s snowing lemma saying, roughly, for each Γ -formula $\varphi(x)$ that: the Γ -set coded by φ consists exactly of all numbers x satisfying $\varphi(x)$.

2.52 Theorem. For $n > 0$, each of the theories $I\Sigma_n$, $B\Sigma_{n+1}$, $P\Sigma_n$ is finitely axiomatizable.

Proof (sketch). Let n be given. We show that $I\Sigma_n$ is finitely axiomatizable. Observe that the assertion " $Sat_{\Sigma, n}$ satisfies Tarski truth conditions for Σ_n -formulas" is a conjunction of finitely many conditions (one for atomic formulas, one for each connective, two for the bounded quantifiers and two for unbounded ones, say). Thus it is a single formula provable in $I\Sigma_1$; let I_n be the finite subtheory of $I\Sigma_1$ making this formula meaningful and provable. In I_n we can express the single sentence saying

$$(*) \quad (\forall X \ \Sigma_n^{\bullet}\text{-set}^{\bullet}) \\ [(0 \in_{\Sigma, n} X \ \& \ (\forall x)(x \in_{\Sigma, n} X \rightarrow S(x) \in_{\Sigma, n} X)) \rightarrow (\forall x)(x \in_{\Sigma, n} X)]$$

(each $\Sigma_n^{\bullet}$ -set satisfies induction). Observe that this is in fact one particular instance of Σ_n -induction (since $\in_{\Sigma, n}$ is Σ_n , thus $(*)$ is provable in $I\Sigma_n$; on the other hand, each instance of Σ_n -induction follows from $I_n + (*)$). Indeed, take a Σ_n -formula $\sigma(x, y)$ (y being a parameter); then $\bar{\varphi}' = Subst^{\bullet}(\bar{\varphi}, \bar{y}, \bar{y})$ is a $\Sigma_n^{\bullet}$ -set – a formula[•] with one free variable. Adding possibly one new axiom we may prove $Sat_n(\bar{\varphi}, \{(\bar{x}, \bar{x}), (\bar{y}, \bar{y})\}) \equiv Sat_n^{\bullet}(\bar{\varphi}', \{(\bar{x}, \bar{x})\}) \equiv \varphi(x, y)$. Now the class X coded by $\bar{\varphi}'$ is inductive by $(*)$; this gives I_{φ} . $\square$

2.53 Definition ($I\Sigma_1$). A set X is *piecewise coded* (p.c.), if for each u there is a sequence s of zeros and ones of length u such that $(\forall i < u)((s)_i = 1 \equiv i \in X)$.

Remark. This is in fact a *scheme* of definitions, depending on the chosen notion of a set and membership (cf. 2.51). Recall also Def. 2.6 (a formula is piecewise coded in a theory $T \supseteq I\Sigma_1$). Evidently, the relation is as follows: $\varphi(x_0, \dots, x_n)$ is p.c. in T iff T proves that the set of all n -tuples $\langle x_0, \dots, x_n \rangle$ such that $\varphi(x_0, \dots, x_n)$ is p.c.

2.54 Definition. (1) Let X be a new variable and let new atomic formulas be the old ones plus $t \in X$ where t is any term. $\Sigma_0(X)$ formulas result from new atomic formulas using connectives and bounded quantifiers.

(2) Copy the definition in $I\Sigma_1$, i.e., define $\Sigma_0^{\bullet}(X^{\bullet})$ formulas in the obvious way. Clearly, φ is a $\Sigma_0(X)$ -formula iff $I\Sigma_1 \vdash [\bar{\varphi} \text{ is a } \Sigma_0^{\bullet}(X^{\bullet})\text{-formula}]$.

(3) Analogously for $\Sigma_n(X)$, $\Pi_n(X)$, $\Sigma_n^{\bullet}(X^{\bullet})$, $\Pi_n^{\bullet}(X^{\bullet})$.

2.55 Main Theorem. (Satisfaction for $\Sigma_0^{\bullet}(X^{\bullet})$ -formulas.) There is a formula $Sat_{0, X}(z, c)$ such that $I\Sigma_1$ proves the following:

If X is p.c. then $Sat_{0, X}$ obeys Tarski truth conditions for $\Sigma_0^{\bullet}(X^{\bullet})$ formulas with $X^{\bullet}$ interpreted as X . Furthermore, under this interpretation each $\Sigma_0^{\bullet}(X^{\bullet})$ set is p.c. and hence satisfies the least number principle. Under the assumption " X is p.c.", Sat_0 is $\Delta_1(X)$ in $I\Sigma_1$.

2.56 Remark. Clearly, this is again a scheme dependent on which notion of sets and membership is used. We can summarize the theorem by saying that in I_1 we can define satisfaction for sets Σ_0 definable from a set X *provided* X is p.c. Thus for a p.c. set X , we have coding for $\Sigma_0(X)$ sets; they are all p.c. and hence satisfy the least number principle. The enigmatic formulation saying that Sat_0 is $\Delta_1(X)$ under the assumption of X being p.c. means, pedantically, the following: we have a $\Sigma_1(X)$ formula Sat_0 and a $\Pi_1(X)$ formula Sat'_0 and $I\Sigma_1 \vdash [X \text{ p.c.} \rightarrow Sat_{0,X} \equiv Sat'_{0,X}]$.

2.57 Corollary (Satisfaction for $\Sigma_k^*(X^*)$ formulas and $\Pi_k^*(X^*)$ formulas.) For each $k > 0$, there is a $\Sigma_k(X)$ -formula $Sat_{\Sigma,k,X}(z, e)$ such that $I\Sigma_1$ proves the following:

If X is p.c. then $Sat_{\Sigma,k,X}$ obeys Tarski truth conditions for $\Sigma_k^*(X^*)$ -formulas, X^* being interpreted as X .

Similarly for $\Pi_k(X)$.

(Obviously, $Sat_{\Sigma,k,X}$ is constructed from $Sat_{0,X}$ – or more precisely, from the two forms of $Sat_{0,X}$, exactly as $Sat_{\Sigma,k}$ was constructed from Sat_0 .)

Caution. Nothing is claimed on $\Sigma_k^*(X^*)$ -sets being p.c.!

2.58 Corollary (“It’s snowing”-it’s snowing lemma). If $\varphi(x_0, \dots, x_k)$ is $\Sigma_n(X)$ then $I\Sigma_1$ proves the following:

If X is p.c. then $[\varphi(x_0, \dots, x_k) \equiv Sat_{\Sigma,n,X}(\overline{\varphi}(\dot{x}_0(\overline{x}_0, \dots)), \emptyset)]$.

The rest of the subsection elaborates the proof of 2.55 and contains some additional technical devices.

2.58 Definition ($I\Sigma_1$). q is a *partial satisfaction* for $\Sigma_0^*(X^*)$ -formulas^{*} $\leq p$, their evaluations by numbers $\leq r$ and partial interpretation of X by a string s of zeros and ones (in symbols: $PSat'_0(q, p, v, s)$ if q is a finite mapping whose domain consists of all pairs (z, e) where z is $\Sigma_0^*(X^*)$, $z \leq p$, e is an evaluation for z , $e \subseteq (\leq p) \times (\leq v)$, $range(q) \subseteq \{0, 1\}$, $lh(s) > r^p$ and Tarski conditions hold for q whenever defined, i.e. we have the same conditions as in 1.71 and, in addition, (v) if z is $t \in X^*$ where t is a term^{*} then $q(z, e) = 1$ iff $(s)_{Val(t,e)} = 1$).

Remark. (1) Note that general power x^y is definable in $I\Sigma_1$ by the usual inductive conditions, see 1.58.

(2) Prove in $I\Sigma_1$ that if t is a term^{*}, $t \leq p$ and e its evaluation^{*} by numbers $\leq r$ then $Val(t, e) \leq r^p$; this follows from the fact that under our assumption on e , $Val(t, e) \leq r^{lh(t)}$, which is proved easily by induction on $lh(t)$. Thus the condition $lh(s) \geq r^p$ guarantees that $(s)_{Val(t,e)}$ is defined.

2.59 Lemma. (1) $PSat_0^*$ is Δ_1 in $I\Sigma_1$.

(2) $I\Sigma_1$ proves the following: If $PSat_0^*(q, p, r, s)$, $PSat_0^*(q', p', r', s')$ and s coincides with s' on the intersection of their domains then q coincides with q' on the intersection of their domains.

(3) Furthermore, $I\Sigma_1$ proves: for each p, r, s such that s is a string of zeros and ones and $lh(s) \geq r^p$ there is a q such that $PSat_0^*(q, p, r, s)$.

Proof is fully analogous to the proof of 1.72. To get (3) prove the following by induction on p (r and s being parameters):

$$lh(s) \geq r^p \rightarrow (\exists q) PSat_0^*(q, p, r, s). \quad \square$$

2.60 Definition ($I\Sigma_1$). Let X be p.c. Define $Sat_{0,X}(z, e)$ iff there are q, p, r, s , such that s is a piece of X , $PSat_0^*(q, p, r, s)$ and $q(q(z, e) = 1)$.

Note that $Sat_{0,X}(z, e)$ implies that z is a $\Sigma_0^*(X^*)$ -formula and e is its evaluation^{*}.

2.61 Proof of the Main Theorem 2.55. We prove that the formula $Sat_{0,X}(z, e)$ has the properties stated. Work in $I\Sigma_1$. First observe that for $z \in \Sigma_0^*(X^*)$, e its evaluation^{*}, and assuming X to be p.c. (so that we have arbitrarily long pieces of X) Lemma 2.59 implies the following:

$$(*) \quad \neg Sat_{0,X}(z, e) \text{ iff there are } q, p, r, s \text{ such that } s \text{ is a piece of } X, PSat_0^*(q, p, r, s) \text{ and } q(\neg^* z, e) = 1.$$

This shows that $Sat_{0,X}$ obeys Tarski condition for negation. Looking from outside $I\Sigma_1$ observe that the definition of $Sat_{0,X}$ is $\Sigma_1(X)$ in $I\Sigma_1$ and $(*)$ gives a $\Pi_1(X)$ definition of $Sat_{0,x}$ under the assumption that X is p.c. The proof of other Tarski conditions is similar and is left to the reader.

Work again in $I\Sigma_1$ and assume X p.c. Then we may speak on $\Sigma_0^*(X^*)$ sets, X^* being interpreted as X (briefly, speak on $\Sigma_0^*(X)$ sets). It remains to be shown that each $\Sigma_0^*(X)$ set is p.c. But this is now trivial: if $z \in \Sigma_0^*(X^*)$ is a formula^{*} with exactly one free variable^{*} and w is arbitrary, then the w -piece of z is easily obtained from any q such that $PSat_0^*(q, p, r, s)$, where $p > z$, $r > w$ and s is a satisfactorily long piece of X . This completes the proof. $\square$

We close this subsection with a lemma on $\Sigma_1^*(X)$ formulas which will be useful later. Recall that having proved 2.55 we also have $Sat_{\Sigma,n,X}$ and $Sat_{\Pi,n,X}$ (cf. 2.57).

2.62 Lemma. There is a formula $WSat_{\Sigma,1}$ which is Δ_1 in $I\Sigma_1$ and such that $I\Sigma_1$ proves the following: for each X p.c., each $z \in \Sigma_1^*(X^*)$ with exactly one free variable and each x , the following are equivalent:

- (i) $Sat_{\Sigma_1, X}(z, [x])$ (where $[x]$ is the evaluation of the free variable of z by x);
- (ii) $(\exists s \text{ piece of } X) WSat_{\Sigma_1}(z, x, s)$;
- (iii) $(\exists w)(\forall s \text{ piece of } X \text{ longer than } w) WSat_{\Sigma_1}(z, x, s)$. Furthermore $I\Sigma_1$ proves $WSat_{\Sigma_1}(z, x, s) \& s \subseteq s' \rightarrow WSat_{\Sigma_1}(z, x, s')$ (monotonicity).
 $WSat_{\Sigma_1}(z, x, s)$ is read: s witnesses the satisfaction of z by x .

Proof. $WSat_{\Sigma_1}(z, x, s)$ says: there are $u, z_1 \leq z$ such that $z = (\exists^\bullet u)z_1$, $z_1 \in \Sigma_0^\bullet(X^\bullet)$ and there is a $y \leq s$ such that for *some* q, p, r we have: $PSat_0^\bullet(q, p, r, s)$ and $q(z_1, [x, y]) = 1$ (with the obvious meaning of $[x, y]$).

Clearly, this is equivalent to saying that there are $u, z_1 \leq z$ as above and a $y \leq s$ such that for *all* q, p, r such that $PSat_0^\bullet(q, p, r, s)$ and $q(z_1, x, y)$ is defined, we have $q(z_1, x, y) = 1$. The rest is evident. $\square$

(d) Relativized Hierarchy in Fragments

Here we shall investigate sets p.c. in $I\Sigma_n$ and in $B\Sigma_{n+1}$ ($n \geq 1$). In particular, we show that $B\Sigma_{n+1}$ proves all Σ_{n+1} sets to be p.c.; we introduce *low* Δ_{n+1} sets in $B\Sigma_{n+1}$ (in 2.69) and prove their basic properties (2.71). The concept of a low Δ_{n+1} set plays a very prominent role in the next section in connection with the analysis of provable forms of König's lemma. Finally we exhibit a class of sets called $\Sigma_0^{exp}(\Sigma_n)$ sets meaningful in $I\Sigma_n$ (2.73); we show that $I\Sigma_n$ proves each $\Sigma_0^{exp}(\Sigma_n)$ class to be p.c. and Δ_{n+1} . This will be useful for generalizing results on König's lemma proved in $B\Sigma_{n+1}$ to results in $I\Sigma_n$.

2.63 Lemma. (1) $I\Sigma_n$ proves each $\Sigma_n^\bullet$ set to be p.c. (2) $I\Sigma_n$ proves that $\Delta_n^\bullet$ total functions are closed under primitive recursion. (3) $B\Sigma_{n+1}$ proves each $\Delta_{n+1}^\bullet$ set to be p.c.

Proof. Imitate the proof of 2.13 (a) (but now working *inside* $I\Sigma_1$: we just prove one theorem with a universal quantifier over all $\Sigma_n^\bullet$ sets): Let X be a $\Sigma_n^\bullet$ set and let z be given: the z -piece of X is the least sequence s of length $z + 1$ such that $(\forall i \leq z)(i \in X \rightarrow (s)_i = 1)$. (Alternatively, you could derive our assertion directly from 2.13 (a) using the fact that the formula Sat_{Σ_1} is Σ_1 and therefore piecewise coded in $I\Sigma_1$.)

(2) Routine.

(3) Imitate the proof of 2.19 inside $B\Sigma_{n+1}$. $\square$

2.64 Definition ($I\Sigma_1$). (1) A set X is *unbounded* if $(\forall x)(\exists y > x)(y \in X)$, i.e. $(Cx)(x \in X)$.

(2) A set X has the *order-type of the universe* (o.t.u.) if for each x there is a sequence s of length x enumerating increasingly the first x elements s , i.e.

$$Seq(s) \& lh(s) = x \& (\forall i < x)((s)_i \in X) \& (\forall i < j < x)((s)_i < (s)_j) \& \\ (\forall i < x)(\forall y < (s)_i)(y \in X \rightarrow (\exists j < i)(y = (s)_j)).$$

2.65 Lemma. For $n \geq 1$, $I\Sigma_n$ proves that each $\Delta_n^\bullet$ unbounded set has o.t.u.

Proof. If X is $\Delta_n^\bullet$ unbounded then the function $F(x) = (\text{least } y)(y \in X \ \& \ y > x)$ is $\Delta_n^\bullet$ total; thus the function G defined by primitive recursion to be

$$\begin{aligned} G(0) &= \min X \\ G(x+1) &= F(G(x)) \end{aligned}$$

is $\Delta_n^\bullet$, total and G is the increasing enumeration of *all* elements of X . G is p.c.; thus for each x , the restriction of G to $(\leq x)$ is a set – the desired sequence s . $\square$

2.66 Corollary (of 2.63). (1) In $I\Sigma_n$, we have satisfaction for $\Sigma_0^\bullet(\Sigma_n^\bullet)$ formulas and therefore may quantify over $\Sigma_0^\bullet(\Sigma_n^\bullet)$ sets. $I\Sigma_n$ proves that each $\Sigma_0^\bullet(\Sigma_n^\bullet)$ set is p.c.

(2) In $B\Sigma_{n+1}$, we have satisfaction for $\Sigma_0^\bullet(\Delta_{n+1}^\bullet)$ and therefore may quantify over $\Sigma_0^\bullet(\Delta_{n+1}^\bullet)$ sets. $B\Sigma_{n+1}$ proves that each $\Sigma_0^\bullet(\Delta_{n+1}^\bullet)$ set is p.c.

Remark. Compare (1) with 2.7 (1) or 2.13 (b): there we had a schema, here a single statement. A schema analogous to our present (2) is possible but cumbersome.

2.67 Theorem. ($I\Sigma_1 + B\Sigma_m, m \geq 1$). (1) Each $\Sigma_0^\bullet(\Delta_m^\bullet)$ set is a $\Delta_m^\bullet$ set; thus $\Sigma_0^\bullet(\Delta_m^\bullet)$ sets coincide with $\Delta_m^\bullet$ sets.

(2) $\Sigma_1^\bullet(\Delta_m^\bullet)$ sets coincide with $\Sigma_m^\bullet$ sets.

Proof. A quick way to prove this is to find (actual) formulas $\sigma(x, \text{par}) \in \Sigma_m$ and $\pi(x, \text{par}) \in \Pi_m$ and to show in $I\Sigma_1 + B\Sigma_m$ that if X is $\Sigma_0^\bullet(\Delta_m^\bullet)$ then for some par , $x \in X$ is equivalent both to $\sigma(x, \text{par})$ and to $\pi(x, \text{par})$. The result follows by the “it’s snowing”-it’s snowing lemma. Thus assume X to be $\Sigma_0^\bullet(Y)$ where Y is $\Delta_m^\bullet$; let z be the $\Sigma_0^\bullet(X')$ -formula defining X from Y and let $u \in \Sigma_m^\bullet$, $v \in \Pi_m^\bullet$, both defining Y . Observe that the formula “ s is a piece of Y ” can be written both as $\sigma_1(s, u, v)$, $\sigma_1 \in \Sigma_m$ in $B\Sigma_m$ and $\pi_1(s, u, v)$, $\pi_1 \in \Pi_m$ in $B\Sigma_m$, e.g. σ_1 is

$$(\forall i < lh(s))((s)_i = 1 \rightarrow \text{Sat}_{\Sigma, m}((u, [i]) \ \& \ (s)_i = 0 \rightarrow \neg \text{Sat}_{\Pi, m}(v, i)).$$

Now

$$x \in X \equiv (\exists q, p, r, s)(s \text{ piece of } Y \ \& \ \text{PSat}(q, p, r, s) \ \& \ q(z, [x]) = 1)$$

and

$$x \notin X \equiv (\exists q, p, r, s)(s \text{ piece of } Y \ \& \ \text{PSat}(q, p, r, s) \ \& \ q(z, [x]) = 0).$$

This completes the proof of (1).

(2) is easy: by (1), a $\Sigma_1^\bullet(\Delta_m^\bullet)$ set can be defined by a formula of the form $(\exists^\bullet u)z$, where z is $\Delta_m^\bullet$; in the usual manner $(\exists^\bullet u)z$ may be replaced by a $\Sigma_1^\bullet$ -formula (contraction of quantifiers). $\square$

2.68 Corollary. For $n, m \geq 1$, $B\Sigma_{n+m}$ proves that each $\Sigma_{n+1}^\bullet(\Delta_m^\bullet)$ set satisfies the least number principle and is p.c.

2.69 Definition ($B\Sigma_{n+1}$, $n \geq 1$). A set X is *low* $\Delta_{n+1}^\bullet$ if it is $\Delta_{n+1}^\bullet$ and each $\Sigma_1^\bullet((X))$ set is also $\Delta_{n+1}^\bullet$.

Remark. This notion comes from recursion theory. Note that sets Δ_2 definable in $\mathbb{N}$ are exactly all sets recursive in K (the Σ_1 -complete set), i.e. sets X such that $dg(X) \leq 0'$ (dg is the Turing degree). A Δ_2 set X is *low* if $(dg((X)))' = 0'$, i.e. the jump of the degree of X is as small as possible. A reader not familiar with the notions involved may disregard this remark.

2.70 Remark. The definition of a low $\Delta_{n+1}^\bullet$ set is meaningful in $B\Sigma_{n+1}$ ($n \geq 1$) since this theory has satisfaction for $\Sigma_1^\bullet(\Delta_{n+1}^\bullet)$ formulas. Obviously, in $B\Sigma_{n+1}$ we may quantify over low $\Delta_{n+1}^\bullet$ sets: they are just some particular $\Delta_{n+1}^\bullet$ sets. Let us generalize: we may speak on low $\Delta_k^\bullet(X)$ sets if we know that $\Delta_k^\bullet(X)$ sets are p.c.

2.71 Theorem. The following equalities are expressible and provable in $B\Sigma_1(m \geq 2)$:

$$\begin{aligned} \text{low } \Delta_m^\bullet &= \text{low } \Delta_2^\bullet(\text{low } \Delta_m^\bullet) = \text{low } \Delta_2^\bullet(\Delta_{m-1}^\bullet) \\ &= \Delta_1^\bullet(\text{low } \Delta_m^\bullet), \quad \Sigma_2^\bullet(\text{low } \Delta_m^\bullet) = \Sigma_m^\bullet \end{aligned}$$

Proof. Work in $B\Sigma_m$. We know that low $\Delta_m^\bullet$ sets are p.c. (since $\Delta_m^\bullet$ sets are); therefore each set $\Sigma_0^\bullet$ -defined from a low $\Delta_m^\bullet$ set X is p.c., thus also $\Delta_k^\bullet(X)$ makes sense. Moreover, if X is low $\Delta_m^\bullet$ and Y is $\Delta_2^\bullet(X)$ then Y is $\Delta_m^\bullet$ (since it is $\Sigma_1^\bullet$ in a $\Pi_1^\bullet(X)$ set, i.e. $\Sigma_1^\bullet$ in a $\Delta_m^\bullet$ -set, i.e. $\Sigma_m^\bullet$, and similarly, Y is $\Pi_m^\bullet$) and thus Y is p.c. We have proved $\Delta_2^\bullet(\text{low } \Delta_m^\bullet) \subseteq \Delta_m^\bullet$. But then $\text{low } \Delta_2^\bullet(\text{low } \Delta_m^\bullet) \subseteq \text{low } \Delta_m^\bullet$. This completes the proof. $\square$

*

In the rest of this subsection we discuss $I\Sigma_n$ instead of $B\Sigma_{n+1}$.

2.72 Theorem ($I\Sigma_n$, $n \geq 1$). Each $\Sigma_0^\bullet(\Sigma_n^\bullet)$ set is a $\Delta_{n+1}^\bullet$ set.

Proof. Similar to 2.67, we find formulas $\sigma(x, c, z)$ and $\pi(x, c, z)$ Σ_{n+1} and Π_{n+1} in $I\Sigma_n$ such that $I\Sigma_n$ proves the following:

If X_c is $\Sigma_n^\bullet$ -defined by $c \in \Sigma_n^\bullet$ and Y_z is $\Sigma_n^\bullet(X^\bullet)$ -defined by z from X_c then $(\forall x)(x \in Y \equiv \sigma(x, c, z) \equiv \pi(x, c, z))$.

By the "it's snowing"-it's snowing lemma we get that Y_z is a $\Delta_{n+1}^\bullet$ set. Details follow.

Now, $I\Sigma_n$ proves

$$\begin{aligned}
 (*) \quad x \in Y_z &\equiv (\exists)(s \text{ is a piece of } X_c \ \& \ (\exists q, u, v)(PSat(q, u, s) \\
 &\quad \& \ q(z, [x]) = 1) \\
 &\equiv (\forall s)(s \text{ is a piece of } X_c \rightarrow (\forall q, u, v)(PSat((q, u, s) \\
 &\quad \& \ q(z[x]) \text{ defined} \rightarrow q(z[x]) = 1)).
 \end{aligned}$$

Here " s is a piece of X " is $\Sigma_0(Sat_{\Sigma, n})$, therefore $\Sigma_0(\Sigma_n)$ in $I\Sigma_n$; by 2.25, it is Δ_{n+1} in $I\Sigma_n$. Hence the formula $(\exists s)(\dots)$ in $(*)$ is Σ_{n+1} in $I\Sigma_n$ and so $(\forall s)(\dots)$ is Π_{n+1} in $I\Sigma_n$. This completes the proof. $\square$

2.73 Definition (cf. 2.54). (1) Let X be a new variable and add $t \in X$ (t a term) to the atomic formulas. $\Sigma_0^{exp}(X)$ formulas result from new atomic formulas using connectives and bounded quantifiers of the form $(\forall x \leq y)$, $(\forall x \leq 2^y)$ and similarly for $\exists$.

(2) In $I\Sigma_1$ define $\Sigma_0^{exp}(X^\bullet)$ formulas copying the definition.

2.74 Theorem (cf. 2.55). There is a formula $Sat_{0, X}^{exp}(z, e)$ such that $I\Sigma_1$ proves the following:

If X is p.c., then $Sat_{0, X}^{exp}$ obeys Tarski truth conditions for $\Sigma_0^{exp}(X^\bullet)$ formulas with $X^\bullet$ interpreted as X , and each $\Sigma_0^{exp}(X)$ set is p.c. and thus satisfies the least number principle. Under the assumption " X is p.c." $Sat_{0, X}^{exp}$ is $\Delta_1(X)$ in $I\Sigma_1$.

Proof. We shall define $PSat_0^{exp}(q, p, r, s)$ (partial satisfaction) analogously to 2.58, but we must be sure that if q is defined for a formula $(\forall u < 2^v)\varphi(u, \dots)$ and an e evaluating v by some x then q will be defined for the evaluation e' extending e and evaluating u by any number $\leq 2^x$. To achieve this make the following definitions:

$qd(z)$ is the quantifier depth of a formula: $qd(z) = 0$ for open formulas, $qd(z_1 \& z_2) = \max(qd(z_1), qd(z_2))$, $qd(\neg z) = qd(z)$, $qd((\forall u \leq 2^v)z) = qd(z) + 1$.

2_y^x is the iterated power of 2: $2_0^x = x$, $2_{y+1}^x = 2^{2_y^x}$.

Then require in the definition of $PSat_0^{exp}$ that $q(z, e)$ be defined whenever $z \leq u$ and $2_{qd(z)}^{\max(e)} \leq v$; the rest is the same as in 2.58, in particular, $v^u \leq lh(s)$ is assumed. Given this, it is easy to prove the following:

- $PSat_0^{exp}$ is Δ_1 in $I\Sigma_1$,
- if $PSat_0^{exp}(q_i, u_i, v_i, s_i)$ ($i = 1, 2$) and s_1 coincides with s_2 where defined then q_1 coincides with q_2 where defined;

- for each u, v and each string s such that $lh(s) > v^u$, there is a q such that $PSat_0^{exp}(q, u, v, s)$.

Then $Sat_{0,X}^{exp}(z, e)$ is defined as

$$(\exists q, u, v, s) \quad (s \text{ is a piece of } X, PSat_0^{exp}(q, u, v, s) \text{ and } q(z, e) = 1);$$

assuming X p.c., $Sat_{0,X}^{exp}(z, e)$ is equivalent to $(\forall q, u, v, s) (s \text{ is a piece of } X \text{ and } PSat_0^{exp}(q, u, v, s) \text{ and } q(z, e) \text{ defined} \rightarrow q(z, e) = 1)$. $\square$

2.75 Corollary. We may introduce $Sat_{\Sigma, n, X}^{exp}$ in the obvious way and for each n prove in $I\Sigma_1$ the following: if X is p.c. then $Sat_{\Sigma, n, X}^{exp}$ and $Sat_{\Pi, n, X}^{exp}$ obey the respective Tarski truth condition.

Consequently, we have the corresponding “it’s snowing”-it’s snowing lemma.

2.76 Theorem. $I\Sigma_n$ proves that each $\Sigma_0^{exp}(\Sigma_n)$ set is p.c. and is Δ_{n+1} .

The proof of the fact that each Σ_0^{exp} set is p.c. is routine and uses the finite partial satisfactions. To prove that each $\Sigma_0^{exp}(\Sigma_n)$ set is Δ_{n+1} , first show that each (actual) $\Sigma_0^{exp}(\Sigma_n)$ formula is Δ_{n+1} in $I\Sigma_n$ (generalizing 2.49–2.50) and then imitate the proof of 2.72.

2.77 Definition. (1) In the above exponentiation may be replaced by any total Δ_1 function H ; moreover, in $I\Sigma_n$ we may define X to be a $\Sigma_0^*(\Sigma_n)$ set if for some total Δ_1 function H and some Σ_n set Y , X is $\Sigma_0^H(X)$. We have satisfaction for $\Sigma_0^*(\Sigma_n)$ sets obeying Tarski truth conditions.

(2) We may define low $\Sigma_0^{exp}(\Sigma_n)$ sets (or low $\Sigma_0^*(\Sigma_n)$ sets) in $I\Sigma_n$: a set X is low $\Sigma_0^H(\Sigma_n)$ if it is $\Sigma_0^H(\Sigma_n)$ and each $\Sigma_1(X)$ set is $\Sigma_0^H(\Sigma_n)$. This is useful for generalizations of the low basis theorem in the next section.

(3) From here on out we shall write LL_n instead of low $\Sigma_0^*(\Sigma_n)$ (n -very low sets).

2.78 Lemma. For $n \geq 1$, $I\Sigma_n$ proves $I\Sigma_1(LL_n)$ and $B\Sigma_1(LL_n)$, i.e. induction and collection for $\Sigma_1(LL_n)$ sets.

Proof. This follows from the obvious modification of 2.76 (for $\Sigma_0^*(\Sigma_n)$ sets rather than $\Sigma_0^{exp}(\Sigma_n)$ sets) and the fact that each $\Sigma_1(LL_n)$ set is a $\Sigma_0^*(\Sigma_n)$ set: we get induction. Collection follows from induction in the usual way. $\square$

Caution. In saying “ X is $\Sigma_1(LL_n)$ ” we mean that for some LL_n set Y , X is $\Sigma_1(Y)$.

2.79 Theorem. For $n \geq 1$, $I\Sigma_n$ proves $\Delta_1(LL_n) = LL_n$, i.e. if Z is LL_n and $Y \in \Delta_1(Z)$ then Y is LL_n .

Proof. Let Y, Z be as above and let $X \in \Sigma_1(Y)$. Then, for appropriate $\Sigma_1(Z)$ -formulas^{*} φ, ψ , we have

$$\begin{aligned} Y &= \{y \mid (\exists s \text{ piece of } Z) WSat_{\Sigma,1}(\varphi, y, s)\}, \\ -Y &= \{y \mid (\exists s \text{ piece of } Z) WSat_{\Sigma,1}(\psi, y, s)\}. \end{aligned}$$

(For $WSat$ see 2.62.) Using $B\Sigma_1(LL_n)$ we get a common bound:

$$(\forall y < a)(\exists s \text{ piece of } Z)(WSat_{\Sigma,1}(\varphi, y, s) \vee WSat_{\Sigma,1}(\psi, y, s)).$$

Thus, for some Δ_1 formula δ ,

$$t \text{ is a piece of } Y \equiv (\exists s \text{ piece of } Z)\delta(t, s).$$

Now $X \in \Sigma_1(Y)$; i.e. for some $\rho \in \Sigma_1(Y)$,

$$\begin{aligned} X &= \{x \mid (\exists t \text{ piece of } Y) WSat(\rho, x, t)\} = \\ &= \{x \mid (\exists s \text{ piece of } Z)(\exists t)(\delta(t, s) \& WSat_{\Sigma,1}(\rho, x, t))\}. \end{aligned}$$

This shows that $X \in \Sigma_1(Z)$. Thus $\Sigma_1(Y) \subseteq \Sigma_1(Z) \subseteq \Sigma_0^*(\Sigma_n)$ and consequently $Y \in LL_n$. We have shown $\Delta_1(LL_n) \subseteq LL_n$. $\square$

2.80 Corollary. ($I\Sigma_n$). If $Z \in LL_n$ then $\Delta_1(\Delta_1(Z)) = \Delta_1(Z)$.

(e) Axiomatic Systems of Arithmetic with No Function Symbols

There are various situations in which it is useful to work with fragments of arithmetic formalized in a language having no function symbols. We shall encounter such situations repeatedly in this book. In the present subsection we prepare the necessary formalism.

2.81 Definition. L' is the language with a constant 0 (zero), binary predicates $=, \leq, S$ (equality, less-than, successor), ternary predicates A, M (addition, multiplication). Bounded quantifiers are defined in the obvious way; formulas of L' with all quantifiers bounded are called *bounded' formulas* or Σ_0' -formulas. Σ_n' and Π_n' formulas result from Σ_0' -formulas in the obvious way.

2.82 Remark. We describe a rather weak axiom system called BA' . The main idea is that S, A, M may describe partial functions and that there may be a largest element (top). The functions defined by S, A, M have to satisfy the usual inductive conditions whenever the values in question are defined.

Since we do not want to spend much time polishing the axiom system we shall make no optimization (minimalization) of the number of axioms.

2.83 Definition. BA' is the theory in L' having the following axioms:

- (1) Axioms saying that $\leq$ is a discrete linear order with the least element 0.
- (2) $S(x, y)$ iff y is the upper neighbour of x w.r.t. $\leq$.

Further axioms say that A and M define binary operations, possibly partial.

- (3) Induction properties of A and M :

$$A(x, 0, z) \equiv z = x$$

$$S(y, y') \& S(z, z') \rightarrow (A(x, y, z) \equiv A(x, y', z'))$$

$$M(x, 0, z) \equiv z = 0$$

$$S(y, y') \& A(z, x, z') \rightarrow (M(x, y, z) \equiv M(x, y', z'))$$

- (4) Commutativity and associativity of A and M , distributivity, monotonicity of addition, monotonicity of multiplication by a positive number, the relation of $\leq$ to addition: $x \leq y \equiv (\exists u \leq y)A(x, u, y)$.

Caution: equalities are understood to be saying "if one side is defined then the second is too and both sides are equal", e.g.

$$A(x, y, z) \equiv A(y, x, z),$$

$$(\exists u)(A(x, y, u) \& A(u, z, w)) \equiv (\exists v)(A(y, z, v) \& A(x, v, w))$$

(associativity); monotonicity for A reads

$$(A(x, z, u) \& A(y, z, v)) \rightarrow (x \leq y \equiv u \leq v).$$

- (5) Schema of induction for Σ'_0 -formulas:

$$(\varphi(0) \& (\forall x, y)(\varphi(x) \& S(x, y) \rightarrow \varphi(y))) \rightarrow (\forall x)\varphi(x)$$

2.84 Remark. (1) The reader may try to get rid of (parts of) (4) by proving some of these axioms from the remaining ones in analogy to the corresponding proofs in I_{open} .

(2) Show that BA' proves the least number principle in the usual way.

(3) Prove the following in BA' :

$$(A(x, y, z) \& y_0 \leq y) \rightarrow (\exists z_0 \leq z)A(x, y_0, z_0)$$

$$A(x, y, z) \rightarrow (x \leq z \& y \leq z)$$

$$(M(x, y, z) \& y_0 \leq y) \rightarrow (\exists z_0 \leq z)M(x, y_0, z_0)$$

$$(M(x, y, z) \& z \neq 0) \rightarrow (x \leq z \& y \leq z)$$

(or just accept these formulas as further axioms).

2.85 Definition. (a) $I\Sigma'_0$ results from BA' by adding the axiom " S, A, M define total functions" $((\forall x)(\exists y)S(x, y)$ etc.). It follows that there is no top element $((\forall x)(\exists y)(x < y))$. $I\Sigma'_n$ results from $I\Sigma'_0$ by allowing any Σ'_n -formula φ in the induction schema; similarly $I\Pi'_n$.

(b) TA' is arithmetic with a top – the extension of BA' by the axiom "there is a top element" $((\exists x)(\forall y)(y \leq x))$.

2.86 Lemma. TA' proves induction for each L' -formula.

Proof. This is evident: each quantifier can be bounded by the top. □

2.87 Remark. Clearly, in $I\Sigma_0$ we may define S, A, M in the obvious way ($S(x, y) \equiv y = S(x)$, $A(x, y, z) \equiv z = x + y$, etc.); then we obviously get $I\Sigma_0 \vdash I\Sigma'_0$ and similarly $I\Sigma_n \vdash I\Sigma'_n$. On the other hand, in $I\Sigma'_0$ we have axioms stating that S, A, M define total functions, thus we may introduce the operations $S, +, *$. To prove $I\Sigma'_0 \vdash I\Sigma_0$, etc., we need the following.

2.88 Lemma. For each bounded formula $\varphi(\mathbf{x})$ there is a bounded' formula $\varphi'(\mathbf{x}, y)$ (with one new free variable) and a constant k such that $I\Sigma'_0$ (enriched by the function symbols $S, +, *$) proves

$$\bigwedge_i y > x_i^{\bar{k}} \rightarrow \varphi(\mathbf{x}) \equiv \varphi'(\mathbf{x}, y).$$

Proof. The only problem is with atomic formulas and this leads us to terms. For each term t and new variables y, z let $z =_y t$ be defined as the following example shows:

$$z =_y (x_1 + x_2) * x_3 \text{ is} \\ (\exists w_1 \leq y)(\exists w_2 \leq y)(A(x_1, x_2, w_1) \& M(w_1, x_3, w_2) \& w_2 = z);$$

then $z =_y t$ is bounded' and if k is the term t understood to be a natural number then

$$I\Sigma'_0 \vdash y \geq (\max \mathbf{x})^{\bar{k}} \rightarrow (z = t \equiv z =_y t).$$

(Observe that if t consists of n symbols then $k \geq n$; and if $\mathbf{x}$ is the tuple of variables in t then $I\Sigma' \vdash t \leq (\max \mathbf{x})^n$.)

Let $t_1 = t_2$ be atomic and let k be this formula interpreted as a number. Then

$$I\Sigma'_0 \vdash (\max \mathbf{x})^k \leq y \rightarrow (t_1 = t_2 \equiv (\exists z \leq y)(z =_y t_1 \& z =_y t_2))$$

and similarly for $t_1 \leq t_2$. This shows the constructions of φ' for φ atomic. The induction step is easy. □

2.89 Theorem. For each n , $I\Sigma'_n \vdash I\Sigma_n$.

Proof. By the preceding, for any Σ'_0 -formula φ and the corresponding φ' , we have

$$\begin{aligned} I\Sigma'_0 \vdash \varphi(\mathbf{x}) &\equiv (\forall y)(\bigwedge_i x_i^k < y \rightarrow \varphi(\mathbf{x}) = \varphi'(\mathbf{x}, y)) \\ &\equiv (\exists y)(\bigwedge_i x_i^k < y \ \& \ \varphi(\mathbf{x}) \equiv \varphi'(\mathbf{x}, y)). \end{aligned}$$

In $I\Sigma'_0$, assume $\varphi(x, \dots)$; we find a least x such that $\varphi(x, \dots)$; this will show $L\Sigma_0$ and thus $I\Sigma_0$. Let $x, \dots$ be given and take a y such that $(x^k < y \ \& \ \dots)$; then we get $\varphi'(x, \dots, y)$. By $L\Sigma'_0$ (provable in $I\Sigma'_0$) we get a least x such that $\varphi'(x, \dots, y)$; by the above equivalence, this is the least x such that $\varphi(x, \dots)$.

For $I\Sigma'_n$ ($n > 0$) reason similarly; by the above equivalence (and by contraction of quantifiers), each Σ_n -formula is Σ'_n in $I\Sigma'_n$. $\square$

2.90 Corollary. Since all of the reasoning above is formalizable in $I\Sigma_1$ we get

$$I\Sigma_1 \vdash \text{Con}^*(I\Sigma_n^*) \equiv \text{Con}^*(I\Sigma_n'^*)$$

for each n .

3. Fragments and Recursion Theory

Introduction. In this section we shall first prove in $I\Sigma_1$ two advanced theorems concerning the arithmetical hierarchy: the *limit theorem* for Δ_2 functions (subsection (a)) and the *low basis theorem*, which is an effective version of König's lemma (subsection (b)). The classical König's lemma says that an infinite (countable) finitely branching tree has an infinite branch; in arithmetic, we have to assume something about the arithmetical complexity of the tree and the conclusion is the existence of an infinite branch of a certain complexity. (See below for details.) Both the limit theorem and the low basis theorem are heavily used in Chap. 2. In subsection (c) we elucidate the status of the theorem saying that each infinite Σ_1 set contains an infinite Δ_1 subset. Finally, in subsection (d) we formulate Matiyasevič's theorem in $I\Sigma_1$.

(a) Limit Theorem

3.1 Definition ($I\Sigma_1$). Let F be a total function; c is the *limit* of F (notation: $c = \lim_s F(s)$) if $(\exists t)(\forall s \geq t)(F(s) = c)$. Similarly for $G(x) = \lim_s F(x, s)$, etc.

3.2 Uniform Limit Theorem for Functions. $I\Sigma_1$ proves the following: there is a total Δ_1 function F of three arguments such that for each total Δ_2 function G of one argument we have

$$(\forall x)(G(x) = \lim_s (F(e, x, s))),$$

where e is a code of G (i.e. a code of the Δ_2 relation $y = G(x)$).

3.3 Corollary ($I\Sigma_1$). (1) For each Δ_2 total function G of one argument there is a Δ_1 total function F of two arguments such that $(\forall x)(G(x) = \lim_s F(x, s))$.

(2) In particular, for each Δ_2 set X there is a total binary Δ_1 function F such that for each x ,

$$x \in X \equiv \lim_s F(x, s) = 1 \text{ and } x \notin X \equiv \lim_s F(x, s) = 0.$$

3.4 Proof - Part 1. We first deal with Δ_2 sets. We proceed in $I\Sigma_1$. Let X be Δ_2 and let

$$\begin{aligned} x \in X &\equiv (\exists u)(\forall v)\varphi(x, u, v, e) \\ x \in X &\equiv (\forall u)(\exists v)\psi(x, u, v, e) \end{aligned}$$

where $\varphi(x, u, v, e) \equiv \text{Sat}_0(e_0, [x, u, v])$, $\psi(x, u, v, e) \equiv \text{Sat}_0(e_1, [x, u, v])$ and $e = (e_0, e_1)$; e naturally codes X and φ, ψ are Δ_1 . Define (omitting the argument e throughout)

$$\begin{aligned} F(x, s) = 1 &\equiv (\exists u \leq s)[(\forall v \leq s)\varphi(x, u, v) \& (\forall u' < u)(\exists v \leq s)\psi(x, u', v)], \\ F(x, s) &= 0 \text{ otherwise.} \end{aligned}$$

Clearly, F is Δ_1 and total.

We prove

- (1) $x \in X \rightarrow \lim_s F(x, s) = 1$
- (2) $x \notin X \rightarrow \lim_s F(x, s) = 0.$

Assume $x \in X$ and let u_0 be any number such that $(\forall v)\varphi(x, u_0, v)$; by $B\Sigma_1$, there is an $s_0 \geq u_0$ such that $(\forall u' < u_0)(\exists v < s_0)\psi(x, u_0, v)$. Then for each $s \geq s_0$ we have $F(x, s) = 1$.

Now let $x \notin X$, thus $(\exists u)(\forall v)\neg\psi(x, u, v)$ and $(\forall u)(\exists v)\neg\varphi(x, u, v)$. Let u_0 be such that $(\forall v)\neg\psi(x, u_0, v)$ and let $s_0 \geq u_0$ be such that $(\forall u \leq u_0)(\exists v \leq s_0)\neg\varphi(x, u, v)$. Now let u be arbitrary; we show that for each $s \geq s_0$, u does not witness the formula defining $F(x, s)$ to be equal to 1. If $u \leq u_0$ then we get

$(\exists v \leq s) \neg \varphi(x, u, v)$; and if $u_0 < u$ then we get $(\exists u' < u)(\forall v < u) \neg \varphi(x, u', v)$ (namely $u' = u_0$). Thus (1) & (2) have been proved. $\square$

Proof - Part 2. Now let K be a unary total Δ_2 function; by 3.4 there is a total Δ_1 function G such that

$$y = K(x) \equiv \lim_s G(x, y, s) = 1 \quad (3)$$

$$y \neq K(x) \equiv \lim_s G(x, y, s) = 0. \quad (4)$$

Let $V(s, y, x) = \min\{r \leq s \mid r \geq y \text{ \& } (\forall t)(r \leq t \leq s \rightarrow G(x, y, t) = 1)\}$.

$V(s, y, x) = s + 1$ if the above is undefined.

Note that the interval $(V(s, y, x), s)$ is the longest interval ending with s on which $G(x, y, -)$ is constantly equal to 1.

Put $F(x, s) = y$ if

$$[(\forall y' \leq s)(V(s, y', x) \geq V(s, y, x) \text{ and } (\forall y' < y)(V(s, y', x) > V(s, y, x))].$$

Thus $F(x, s) = y$ means that y is the smallest possible number among $y' \leq s$ having the maximal possible interval $(\dots, s)$ (subinterval of (y', s)) on which $G(x, y, -)$ constantly equals to 1. We claim $K(x) = \lim_s F(x, s)$. Let $K(x) = y_0$ and let $s_0 \geq y_0$ be such that $s \geq s_0$ implies $G(x, y_0, s) = 1$. Furthermore, using $B\Sigma_1$ assume $s_1 \geq s_0$ to be large enough to satisfy

$$(*) \quad (\forall y' \leq s_0)(y' = y \rightarrow (\exists s < s_1)(s > s_0 \text{ \& } G(x, y, s) = 0)).$$

Then for $s > s_1$ we have $F(x, s) = y_0$: $V(s, y_0, x) \leq s_0$ but for all other $y' \leq s$ we have $V(s, y', x) > s_0$ by $(*)$ and for $y' > s_0$ since $V(s, y', x) \geq y'$. This completes the proof. $\square$

3.6 Remark. The proofs 3.4–3.5 are due to Švejdar; observe that they prove the limit theorem as a schema in $B\Sigma_1$ ($I\Sigma_1$ is not necessary).

(b) Low Basis Theorem

3.7 Definition ($I\Sigma_1$). A *tree* is a set T of finite sequences containing with each sequence all its initial segments:

$$Tree(T) \equiv (\forall s \in T)(Seq(s)) \text{ \& } (\forall s, t)(Seq(s) \text{ \& } s \subseteq t \text{ \& } t \in T \rightarrow s \in T).$$

T is *finitely branching* if for each $s \in T$ the set of all upper neighbours t of s (i.e. $s \subseteq t \in T$ and $lh(t) = lh(s) + 1$) is bounded. T is Δ_1 -estimated if

there is a Δ_1 function F such that $(\forall x)(\forall s \in T)(lh(s) = x \rightarrow s \leq F(x))$. (Evidently, a Δ_1 -estimated tree is finitely branching.) A subtree $B \subseteq T$ is a *branch* of T if B is linearly ordered by the relation "being an initial segment of" (i.e. by inclusion).

3.8 Low Basis Theorem ($I\Sigma_1$). Each unbounded Δ_1 tree which is Δ_1 estimated has a low $\Sigma_0^*(\Sigma_1)$ unbounded branch.

3.9 Remark. (1) The theorem will be proved in this subsection; for the notion of low $\Sigma_0^*(\Sigma_1)$ sets (or LL_1 sets) see 2.77–78. In particular, recall that $I\Sigma_1$ proves induction and the least number principle for $\Sigma_0^*(\Sigma_1)$ sets, so that the notion of LL_1 sets makes sense.

(2) Recall 2.65: if T is Δ_1 and unbounded then it is o.t.u. (of the order type of the universe). Moreover, if Z is an LL_1 unbounded branch through a Δ_1 tree then Z is also o.t.u. – cf. 2.78: we have $I\Sigma_1(Z)$ and may relativize 2.65.

(3) Thus if $T \in \Delta_1$ is a Δ_1 -estimated unbounded tree and B is an LL_1 unbounded branch through T then B determines a total function

$$Z(x) = y \equiv (\exists s \leq F(x+1))(s \in B \ \& \ (s)_x = y)$$

(where F is an estimator for T). Clearly, Z is $\Delta_1(B)$ and hence LL_1 by 2.79. Call Z a *branch function*. Conversely, an LL_1 branch function determines an LL_1 branch.

3.10 Corollary. Let $k \geq 1$. (1) $I\Sigma_k$ proves that each $T \in LL_k$ which is an unbounded LL_k -estimated tree has an LL_k unbounded branch. (In particular, each Δ_k , Δ_k -estimated unbounded tree has an LL_k unbounded branch.)

(2) $I\Sigma_{k+1}$ proves that each $T \in LL_k$ which is an unbounded finitely branching tree has an LL_{k+1} unbounded branch.

(3) $B\Sigma_2$ proves that each $T \in \Delta_1$ which is an unbounded Δ_1 -estimated tree has a low Δ_2 unbounded branch.

(4) $B\Sigma_{k+1}$ proves that each $T \in low \Delta_{k+1}$ which is an unbounded $low \Delta_{k+1}$ -estimated tree has a $low \Delta_{k+1}$ unbounded branch.

(5) $B\Sigma_{k+2}$ proves that each $T \in low \Delta_{k+1}$ which is an unbounded finitely branching tree has a $low \Delta_{k+2}$ unbounded branch.

3.11 Remark. (1) follows by relativization since $I\Sigma_k$ proves $I\Sigma_1(LL_k)$ and, furthermore, $I\Sigma_k$ proves $low \Sigma_0^*(\Sigma_1(LL_k)) \subseteq LL_k$. (Easy; first prove $\Sigma_0^*(\Sigma_1(LL_k)) \subseteq \Sigma_0^*(\Sigma_k)$.)

(2) follows from (1) for $k+1$: observe in $I\Sigma_{k+1}$ that a LL_k finitely branching tree is $\Pi_1(T)$ -estimated, thus Δ_{k+1} -estimated and LL_{k+1} -estimated.

(3) follows from 3.8: In $B\Sigma_2$ we know that each LL_1 set is low Δ_2 . Having (3) we get (4) and (5) analogously to (1) and (2).

3.12 Lemma ($I\Sigma_1$). Let T be a Δ_1 , unbounded, Δ_1 -estimated tree; then there is an $s \in T$ of length 1 such that $T_s = \{t \supseteq s \mid t \in T\}$ is unbounded.

Proof. First observe that a Δ_1 -estimated Δ_1 tree is unbounded iff for each x , there is a $t \in T$ of length x (thus being unbounded is Π_1 for such a tree). Without loss of generality we may assume that the elements of T of length 1 are $(0), (1), \dots, (a-1)$; assume that all $T_{\{i\}}$, $i < a$, are bounded, i.e.

$$(\forall i < a)(\exists h)(\forall s < F(h))(lh(s) = h \rightarrow s \notin T_{\{i\}})$$

(where F is a Δ_1 estimator of the tree). By $B\Sigma_1$,

$$(\exists h)(\forall i < a)(\forall s < F(h))(lh(s) = h \rightarrow s \notin T_{\{i\}}),$$

thus T has no elements of length h and therefore is unbounded. $\square$

3.13 Remark. We are now going to prove the low basis theorem. Recall the Δ_1 formula $WSat$ such that $I\Sigma_1$ proves, for each X p.c., each $\Sigma_1^\bullet(X^\bullet)$ formula z with just one variable, and each number x ,

$$Sat_{\Sigma,1,X}(z, [x]) \equiv (\exists s \text{ piece of } X) WSat_{\Sigma,1}(z, x, s);$$

$WSat(z, x, s)$ reads “ s witnesses the satisfaction of z by x ” (see 2.62 for more details; we shall omit the superscripts $\Sigma, 1$). If X is a *function* (i.e. a particular binary relation) such that for each u , the restriction $X \upharpoonright u$ exists as a finite sequence (which implies that X is p.c.) it is convenient to change slightly the notion of witnessing by replacing “piece of” with “restriction of”; thus for the new Δ_1 predicate (which we still call $WSat$) we have in $I\Sigma_1$:

If X is a function as above, $z \in \Sigma_1^\bullet(X^\bullet)$ (one free variable) and x is arbitrary, then the following are equivalent:

$$\begin{aligned} & Sat_{\Sigma,1,X}(z, [x]), \\ & (\exists x)(\exists s = X \upharpoonright x) WSat(z, x, s), \\ & (\exists x)(\forall y \geq x)(\forall s = X \upharpoonright y) WSat(z, x, s). \end{aligned}$$

In this subsection we shall use this modification of $WSat$.

3.14. We start the proof of 3.8. We work in $I\Sigma_1$; the proof is an inspection of a usual recursion-theoretic proof. Let T be a Δ_1 unbounded tree, let F be a Δ_1 estimator for T . For technical reasons assume that F also estimates the full dyadic tree, i.e. for each sequence s of zeros and ones of length e , we have $s < F(e)$. The construction proceeds in steps. In step e , we define two strings s_e and c_e of length e ; s_e will be a piece of the desired branch Z , and c_e information about a truncation of T enforced by previous steps. We

use a Δ_1 enumeration (φ_e, a_e) of all pairs consisting of a $\Sigma_1^*(X^*)$ -formula with one free variable and of a number. In step e we decide whether φ_e will be satisfied by a_e and Z , or not. For each e , let (e) be the finite set of all sequences less than $F(e)$ ($s \in (e) \equiv \text{Seq}(s) \ \& \ s \leq F(e)$). Clearly, $\Sigma_0^*(\Sigma_1)$ sets are closed under the quantification of the form $(\exists s \in (e)), (\forall s \in \text{string}(e))$. Two sequences s, t are called compatible if $s \subseteq t$ or $t \subseteq s$.

For $s, c \in (e)$ let $T(e, s, c) = \{t \in T \mid (t \text{ compatible with } s) \ \& \ (\forall i < e)((c)_i = 1 \rightarrow \neg \text{WSat}(\varphi_i, a_i, t))\}$. Obviously, $T(e, s, c)$ is Δ_1 , and so is $T'(e, s, c)$ where $T'(e, s, c) = \{t \mid t \in T(e, s, c) \ \& \ \neg \text{WSat}(\varphi_e, a_e, t)\}$. Finally, let $\text{cond}(e, c, s', c')$ be the following condition (saying how to extend (s, c) to (s', c')):

$(s, c \in (e)) \ \& \ (s', c' \in (e+1)) \ \& \ (s \subseteq s' \ \& \ c \subseteq c') \ \& \ (\text{Case 1} \vee \text{Case 2})$, where

Case 1. $T'(e, s, c)$ is unbounded, $c' = c \smallfrown \langle 1 \rangle$ (concatenation), and $s' = s \smallfrown \langle i \rangle$, where $i < F(e+1)$ is minimal such that $T'(e, s, c)$ is unbounded over $s \smallfrown \langle i \rangle$ (and $s' = \emptyset$, say, if such an i does not exist);

Case 2. $T'(e, s, c)$ is bounded, $c' = c \smallfrown \langle 0 \rangle$ and $s' = s \smallfrown \langle i \rangle$ where $i < F(e+1)$ is minimal such that $T(e, s, c)$ is unbounded over $s \smallfrown \langle i \rangle$, $s' = \emptyset$ if such an i does not exist. (When we say that a tree T'' is unbounded over t_0 , we mean that $\{t \in T'' \mid t \supseteq t_0\}$ is unbounded).

Observe that cond is $\Sigma_0^*(\Sigma_1)$ (since, as we know, “ T is unbounded” is Π_1 , see above).

3.15 Fact. (1) If $T(e, s, c)$ is unbounded then $(\exists s', c') \text{cond}(e, s, c, s', c')$.

(2) If $T(e, s, c)$ is unbounded and $\text{cond}(e, s, c, s', c')$, then $T(e+1, s', c')$ is unbounded (and contains s').

3.16. Define

$$\text{Path}(e, s, c) \equiv s, c \in (e) \ \& \ (\forall i < e) \text{cond}(i, s \upharpoonright i, c \upharpoonright i, s \upharpoonright (i+1), c \upharpoonright (i+1))$$

(where $s \upharpoonright i$ is the initial segment of s of length i , etc.). Observe that Path is $\Sigma_0^*(\Sigma_1)$.

3.17 Fact. If T is as above, then

$$(\forall e)[(\exists! s, c) \text{Path}(e, s, c) \ \& \ (\forall s, c)(\text{Path}(e, s, c) \rightarrow T(e, s, c) \text{ is unbounded})].$$

(This can be proved by induction, since $[\dots]$ is $\Sigma_0^*(\Sigma_1)$. The induction step follows by 3.15.)

3.18. Define $Z(x) = y \equiv (\exists s, c \in (x+1))(\text{Path}(x+1, s, c) \ \& \ (s)_x = y)$. Clearly, Z is a $\Sigma_0^*(\Sigma_1)$ total function; it is a branch through T .

3.19 Lemma. $\text{Sat}_{\Sigma_1, Z}(\varphi_e, [a_e])$ iff for the unique s, c satisfying $\text{Path}(e, s, c)$ we have Case 2, i.e. $T'(e, s, c)$ is bounded.

Proof. Note that, by 2.22, $Sat_{\Sigma_1, Z}(\varphi_e, a_e)$ is equivalent to the existence of a piece t of Z such that $WSat(\varphi_e, a_e, t)$. Assume Case 2, i.e. $T'(e, s, c)$ is bounded, and let t be a piece of Z longer than a bound for $T'(e, s, c)$. Then $t \in T(e, s, c)$ and $WSat(\varphi_e, a_e, t)$; thus $Sat_{\Sigma_1, Z}(\varphi_e, a_e)$. Now assume Case 1. Then we can prove for each $i > e$

$$(\forall s', c' \in (i))(Path(i, s', c') \rightarrow \neg WSat(\varphi_e, a_e, s'))$$

by induction on i (since the formula in question is $\Sigma_0^*(\Sigma_1)$). Thus there is no piece t of Z such that $WSat(\varphi_e, a_e, t)$. $\square$

3.20. Consequently, Z is low $\Sigma_0^*(\Sigma_1)$; we have

$$Sat_{\Sigma_1, Z}(\varphi_e, a_e) \equiv (\exists s, c \in (e+1) \mid (Path(e, s, c) \& (c)_e = 0)),$$

or, in still more detail, given φ, a , let $e = e(\varphi, a)$ iff $(\varphi, a) = (\varphi_e, a_e)$; then

$$\begin{aligned} Sat_{\Sigma_1, Z}(\varphi, a) &\equiv \\ &\equiv (\exists s, c \in (e(\varphi, a) + 1)) Path(e(\varphi, a), s, c) \& (c)_{e(\varphi, a)} = 0). \end{aligned}$$

This completes the proof of the Low basis theorem.

3.21 Remark. An inspection of the proof gives the following in $I\Sigma_1$: if T is a dyadic unbounded Δ_1 tree (i.e. elements of T are sequences or zeros and ones) then T has a low $\Sigma_0^{exp}(\Sigma_1)$ unbounded tree.

(c) Infinite Δ_1 Subsets

In this subsection we pay attention to the fact that in $B\Sigma_m$ ($m \geq 2$) we have two notions of infinity for Σ_m sets: being unbounded and being o.t.u. (cf. again 2.64–65). We show that $I\Sigma_m$ proves that a Σ_m set is unbounded iff it is o.t.u. ($m \geq 1$); and obviously, $I\Sigma_1$ proves that if a Σ_m set is o.t.u. then it is unbounded. But in general, $B\Sigma_m$ does not prove that each unbounded Σ_m set is o.t.u., thus our notions differ, even for Σ_m sets ($m \geq 2$). The well-known recursion-theoretic fact that an infinite Σ_m set has an infinite Δ_m subset has two formalizations in $B\Sigma_m$ – for our two notions of finiteness. We shall show that both formalizations are provable. (For $m = 1$ we have to add something to make quantification over Σ_1 sets possible – let us just add $I\Sigma_1$).

3.22 Theorem. (a) $I\Sigma_1$ proves that each unbounded Σ_1 set has an unbounded Δ_1 subset.

(b) For $n \geq 1$, $I\Sigma_n$ proves that each unbounded Σ_{n+1} set has an unbounded Δ_{n+1} subset.

Proof. For $n \geq 0$ work in $I\Sigma_{\max(1,n)}$ and let $x \in X \equiv (\exists y)\psi(x, y)$ where ψ is Π_n (for $n = 0$, ψ is Δ_1). For each u , let $F(u)$ be the $v \geq u$ such that $(\exists y)\psi(v, y)$ with the smallest possible witness, i.e.

$$F(u) = v \equiv (\exists w)(\psi(w_0), (w)_1) \& (w)_0 = v \geq u \& \\ \& \neg(\exists w' < w)(\psi((w')_0, (w')_1) \& (w')_0 \geq u).$$

Then F is a total Σ_{n+1} function (due to $L\Pi_n$ or $L\Delta_1$), hence a Δ_{n+1} function, and for $Y = \text{range}(F)$ we have $v \in Y \equiv F(v) = v$ ($\Leftarrow$ is trivial; conversely, if $F(u) = v$ then $F(v) = v$ by the definition of F) and Y is clearly unbounded. Thus Y is a Δ_{n+1} unbounded subset of X . $\square$

3.23 Theorem. For $n \geq 1$, $I\Sigma_n$ proves that each unbounded Σ_n set is o.t.u.

Proof. For Δ_n sets see 2.65. Let X be Σ_n unbounded; recall that X is p.c. By 3.22 and 2.65, let $Y \subseteq X$ be Δ_n and o.t.u. and let s be the sequence of the first x elements of Y . Let a be the maximal element of s and let q be the piece of X up to $a + 1$. Working with q as a finite set and using $I\Sigma_1$ let t be the increasing enumeration of numbers i such that $q(i) = 1$, i.e. $i \in X$ & $i \leq a$. Clearly, each element of the sequence s occurs in t , thus $lh(t) \geq x$. Thus X is o.t.u. $\square$

3.24 Theorem. For $n \geq 1$, $I\Sigma_1 \cup B\Sigma_n$ proves that each Σ_n set which is o.t.u. contains a Δ_n o.t.u. subset.

Proof (a modification of a proof of J. Paris).

Let $x \in A \equiv (\exists y)\theta(x, y)$ where θ is Π_{n-1} (for $n = 0$: θ is Δ_1). Let

$$(u, v) \in R \equiv \theta(u, v) \& (\forall w < v) \neg \theta(u, w),$$

thus R is Δ_n (selector for θ). Consider R as a set of ordered pairs and let $F(a) = b$ if b is the increasing enumeration of the first a elements of R . F is Δ_n ; we show that it is total.

If a is given and c is the sequence coding the first a elements of A then, by $B\Pi_{n-1}$, there is a d such that $(\forall i < a)(\exists y < d)\theta((c)_i, y)$. Thus if e enumerates all elements of R less than $(\max(c), d)$ then e must have at least a members. Thus F is total.

Let $x \in B \equiv (\exists t < x)(\exists v < F(2t + 1))\theta(x, v)$. Then B is Δ_n and $B \subseteq A$. Let a be given; $F(2a + 1)$ is a sequence enumerating the $(2a + 1)$ first elements (u, v) of R . For at most $(a + 1)$ of them we have $u \leq (a + 1)$ (since R is a function); i.e. for at least a of them we have $u > a$ and $v < F(2a + 1)$ and hence $(u, v) \in B$. Thus if s enumerates increasingly all elements of B less than $F(2a + 1)$ then $lh(s) \geq a$. $\square$

(d) Matiyasevič's Theorem in $I\Sigma_1$

Recall Matiyasevič(-Robinson-Davis-Putnam)'s theorem 0.48: it says that Σ_1 subsets (or subrelations) of N coincide with $\exists$ subsets (subrelations), i.e. with subsets defined by purely existential formulas. (Each $\exists$ formula is trivially a Σ_1 formula; thus one inclusion is trivial.) In this short subsection we carefully formulate Matiyasevič's theorem in $I\Sigma_1$. Needless to say, the fact that this important theorem is provable in $I\Sigma_1$ (and even in a weaker theory, see below) is very interesting. But since we shall use this fact only once (in Chap. IV) and since even a proof of the non-formalized Matiyasevič's theorem is rather lengthy but its formalization in $I\Sigma_1$ is more or less immediate, we shall only comment on the proof and shall refer to the literature.

3.25 Theorem. For each Σ_1 formula $\varphi(x_1, \dots, x_n)$ there is a $\exists$ formula $\psi(x_1, \dots, x_n)$ with the same free variables such that $I\Sigma_1 \vdash \varphi \equiv \psi$.

Remark on the proof. Clearly, the theorem is proved by induction on the complexity of φ ; the case of open formulas is obvious and so is the inductive step for $\exists$, $\&$, $\vee$. (For Σ_0 formulas φ one has to prove that both φ and $\neg\varphi$ are equivalent to some $\exists$ formulas.) The only non-trivial, but extremely peculiar, case is that of a bounded universal quantifier – this is the heart of the proof of Matiyasevič's theorem. But realizing this, one can just read an accessible proof (for example [Davis 73], to which we referred also for the proof of 0.84) and check that everything formalizes.

Alternatively, one can consult [Dimitracopoulos 80] where a proof of Matiyasevič's theorem is elaborated even in $I\Sigma_0(exp)$. Detailed instructions for the formalization are also contained in [Gaifman-Dimitracopoulos 82].

Since we have the satisfaction $Sat_{\Sigma,1}$ which is itself a Σ_1 formula, we get the following

3.27 Corollary ($I\Sigma_1$). Each Σ_1^* set is an $\exists^*$ set, i.e. for each Σ_1^* formula z there is an $\exists^*$ formula w (both with just one free variable) such that z and w are equivalent:

$$(\forall x)(Sat_{\Sigma,1}(z, [x]) \equiv Sat_{\Sigma,1}(w, [x])).$$

(Let $\rho(z, x)$ be an $\exists$ formula such that $I\Sigma_1 \vdash Sat_{\Sigma,1}(z, x) \equiv \rho(z, x)$; in $I\Sigma_1$, given z , let w be $Subst^*(\bar{p}, \bar{z}, \dot{z})$, i.e. substitute the numeral^{*} $\dot{z}$ for the variable^{*} $\bar{z}$ in the formula^{*} $\bar{p}$.)

4. Elements of Logic in Fragments

We have already developed some elements of logic in $I\Sigma_1$: in Sect. 1(d) we introduced formulas of arithmetic, $\Sigma_1^\bullet$ formulas, etc., and given k , we defined in $I\Sigma_1$ satisfaction of $\Sigma_k^\bullet$ formulas in the universe of all numbers. Moreover, in Sect. 2 we investigated partial satisfaction for the relativized hierarchy; satisfaction also concerned the universe of all numbers. We shall now extend this in several respects: first, we allow an arbitrary language, not just the language of arithmetic. Second, we shall discuss, inside $I\Sigma_1$, the notion of provability; among other things, we shall be interested in the fact that in $I\Sigma_1$ we have Herbrand's theorem. Third, we shall formalize in $I\Sigma_1$ some facts from general model theory; among other things we prove a version of Gödel's completeness theorem (called the Low arithmetized completeness theorem). Finally we apply these techniques and facts to the language of arithmetic; among other things we show that for each k , $I\Sigma_{k+1}$ proves the consistency of $I\Sigma_k$. A natural continuation is then to formalize, inside $I\Sigma_1$, some facts on non-standard models of arithmetic; but this presupposes that a (informal) model theory of fragments has been elaborated. Thus we shall continue our present investigation of logic inside fragments dealing with models of fragments constructed in fragments in Chap. IV.

In subsections (a), (b) we show how to prove some facts stated in Sect. 0 (0.3–0.21). Note that some model-theoretical proofs do not formalize (or at least their formalization is not known) so that we shall have to use alternative, more elementary proofs.

Note also that a reader hurrying to study Gödel's incompleteness theorems and related topics presented in Chap. III should read subsection (a) of the present section; then he may switch to Chap. III.

(a) Arithmetizing Provability

4.1. We shall work in $I\Sigma_1$ but everything relativizes to $I\Sigma_1(\Gamma)$ where Γ is as in Sect. 3. Recall that in $I\Sigma_1$, total Δ_1 functions are closed under primitive recursion (cf. 1.54). Recall also the theorem on the free algebra of expressions. We now apply this theorem for the chosen construction of terms and formulas of arbitrary Δ_1 language.

4.2 Definition. ($I\Sigma_1$). A Δ_1 language L consists of mutually disjoint Δ_1 sets $\text{Pred}^\bullet$, $\text{Fct}^\bullet$, $\text{Const}^\bullet$ of predicates, functions and constants respectively and a Σ_1 function $\text{ar}^\bullet$ associating with each predicate and function symbol its arity (positive number). We further assume that $\text{Pred}^\bullet$, $\text{Fct}^\bullet$ and $\text{Const}^\bullet$ are disjoint from an infinite Δ_1 set of variables $^\bullet$ and from logical connectives $^\bullet$ and quantifiers $^\bullet$.

The reader can supply without difficulties definitions of the Δ_1 set $\text{Term}^\bullet$ of terms and the Δ_1 set $\text{Form}^\bullet$ of formulas in complete analogy to 1.61 and can prove in $I\Sigma_1$ the properties of formulas and terms analogous to those in 1.61.

Furthermore, we define the Δ_1 set $\text{LogAx}^\bullet$ of logical axioms and two Δ_1 relations-rules of inference, copying the definition in 0.10.

4.3 Definition ($I\Sigma_1$). A *theory* is given by a Δ_1 language L and a set T (not necessarily Δ_1) of formulas of L called *special axioms*, including equality axioms[•] (defined in the obvious way, cf. 0.11). We call T a theory if L is clear from the context. A finite sequence s of L -formulas is a T -proof[•] if for each $i < lh(s)$, $(s)_i$ is a logical or special axiom[•] or follows from some previous members of s by a rule of inference. Note that if T is p.c. then the set $\text{Proof}_T^\bullet$ of all T -proofs is in $\Delta_1(T)$; in particular, if T is Δ_1 then $\text{Proof}_T^\bullet$ is Δ_1 . We write $\text{Proof}_T^\bullet(s, x)$ if s is a T -proof[•] and its last member is x . $\text{Pr}_T^\bullet(x)$ (or $T \vdash^\bullet x$) means that x is T -provable, i.e. $(\exists s)\text{Proof}_T^\bullet(s, x)$.

4.4 Lemma ($I\Sigma_1$ -compactness). If $T \vdash^\bullet x$ then there is a finite subset $t \subseteq T$ such that $t \vdash^\bullet x$.

Proof. (We do not assume $T \in \Delta_1$.) Let $\text{Proof}_T^\bullet(s, x)$ and let t be the set of elements $(s)_i$ of the sequence s such that $(s)_i$ is not a logical axiom and does not follow from previous numbers. Since s is a T -proof, $t \subseteq T$; and $\text{Proof}_t^\bullet(s, x)$. $\square$

4.5 Definition ($I\Sigma_1$). A theory T is Δ_1 -decidable or simply *decidable* if the set of all T -provable formulas is Δ_1 .

4.6 Fact ($I\Sigma_1$). If T is Δ_1 then $\text{Proof}_T^\bullet$ is Δ_1 and $\{x \mid T \vdash^\bullet x\}$ is Σ_1 . (Evident)

4.7 Definition ($I\Sigma_1$). T is *consistent* (in symbols: $\text{Con}_T^\bullet$) if there is no closed formula[•] x such that $T \vdash^\bullet x$ and $T \vdash^\bullet \neg^\bullet x$.

4.8 Remark. A switch to Chap. III is possible now.

4.9 Theorem ($I\Sigma_1$). Let T be a theory and let $T \vdash^\bullet (\exists^\bullet u)\varphi(u)$.

Let c be a constant[•] not in the language of T , let T' be $T + \varphi(c)$ ($\varphi(c)$ is shorthand for $\text{Subst}^\bullet(\varphi, u, c)$). Then T' is a conservative extension of T . If T is decidable then T' is also decidable.

Proof. Copy the usual proof: given a T' -proof, replace each of its members $\alpha(c)$ by $\varphi(x) \rightarrow \alpha(x)$ where x is a new variable. Show by Σ_1 -induction that each $\varphi(x) \rightarrow \alpha(x)$ is T -provable. Thus we get: $T' \vdash^\bullet \alpha(c)$ iff $T \vdash^\bullet \varphi(x) \rightarrow$

$\alpha(x)$ (thus T' is decidable if T is); in particular, if α is a T -formula we get $T' \vdash^\bullet (\exists^\bullet x)\varphi(x) \rightarrow \alpha$ iff $T \vdash^\bullet \alpha$. $\square$

4.10 Theorem ($I\Sigma_1$). Each theory T has a conservative Henkin extension T' , i.e. for each T' -formula $\varphi(u)$ with just one free variable there is a constant c such that $T' \vdash^\bullet (\exists^\bullet u)\varphi(u) \rightarrow^\bullet \varphi(c)$.

If $T \in \Delta_1$ then $T' \in \Delta_1$ and if T is decidable then so is T' .

Proof. This seems to be a trivial consequence of the preceding theorem but there are two difficulties: (1) *First* one has to construct a language L' extending the language of L and a Δ_1 function C associating with each L' -formula φ having just one free variable a constant $C(\varphi)$ (also denoted c_φ) not occurring in φ . (If necessary, we replace L by an isomorphic copy so that we have infinitely many non-members of L). This is easily achieved by generalizing the notion of an expression: take variables and constants of L as atoms and generate simultaneously terms, formulas and constants of L' . (Define a *derivation* of a term, formula or constant in the obvious way and prove all necessary facts). So assume we have L' and C . Extend T by adding all axioms $\varphi(u) \rightarrow^\bullet \varphi(c_\varphi)$: the result is T' . It remains to be proved that T' extends T conservatively. By 4.4 it suffices to show that for each finite $t_1 \subseteq T$ and for each finite set t_2 of the added axioms, $t_1 \cup t_2$ extends t_1 conservatively. We cannot prove this by induction on t_2 since "extends conservatively" is Π_2 and we work in $I\Sigma_1$. This is the *second* obstacle. But assume that p is a $(t_1 \cup t_2)$ -proof of a T -formula φ and prove $t_1 \vdash^\bullet \varphi$.

Let t'_2 be the least subset of t_2 such that $t_1 \cup t'_2 \vdash^\bullet \varphi$ (we use $L\Sigma_1$); if t' is non-empty, 4.9 gives a smaller t'' such that $t_1 \cup t'' \vdash^\bullet \varphi$. Thus t' is empty and $t_1 \vdash^\bullet \varphi$. Moreover, iterated use of 4.9 gives a Δ_1 function D associating with each T' -sentence z a T -sentence $D(z)$ such that $T' \vdash^\bullet z$ iff $T \vdash^\bullet D(z)$; thus if T is decidable then so is T' . $\square$

We have presented a rather detailed proof to show for a relatively simple example some difficulties that we may encounter when formalizing metamathematical proofs in fragments, as well as ways to overcome them. Similar tricks will be used rather frequently in the succeeding chapters.

4.11 Theorem ($I\Sigma_1$). Let L be a language (and assume that its complement is infinite; otherwise take an isomorphic copy).

(1) There is a Δ_1 function PNF associating to each L -formula φ its prenex normal form $\varphi = PNF(\varphi)$ such that ψ is an L -formula, consists of a block of quantifiers followed by an open formula, has the same free variables as φ and the equivalence $(\varphi \equiv \psi)$ is provable in the empty theory with the language L (pure predicate calculus).

(2) There is a language L' extending L by new function symbols and a Δ_1 function F associating with each L' -formula φ and each variable y a function

symbol $F(\varphi, y)$ from $L'-L$ (denoted also by $F_{\varphi, y}$) not occurring in φ , whose arity equals the number of free variables of φ distinct from y .

(3) There is a Δ_1 function Sk associating with each L -formula Φ in prenex normal form its Skolem normal form satisfying conditions copied from 0.15 (where F_i^Φ is $F((Q_{i+1}x_{i+1}) \dots (Q_k x_k)\varphi, x_k)$) in the notation of 0.15, i.e. Φ is $(Q_1 x_1) \dots (Q_k x_k)\varphi(x, y)$, Q_i are quantifiers, $t_i = x_i$ if $Q_i = \forall$, $t_i = F_i^\Phi(\leftarrow t_{i-1})$ if Q_i is $\exists$ and $Sk(\Phi)$ is $\varphi(t_1, \dots, t_k, y)$.

(4) There is a Δ_1 function He associating with each L -formula Φ in prenex normal form its Herbrand normal form, i.e. $He(\Phi)$ is logically equivalent to the existential closure of $\neg Sk(\neg\varphi)$.

Proof. Checking that everything formalizes in IS_1 is trivial except perhaps for (2): but this is achieved in the same way as the construction of Henkin constants in 4.10. $\square$

4.12 Remark. Observe that we have not claimed anything about deductive properties of Skolem and Herbrand normal forms. We shall finally obtain all expected results but now we come to the place where no direct formalization of easy model theoretic proofs is known (at least to the authors, see the problem in 4.28). Thus we have to use a more elementary and tedious approach. We rely on the book by Shoenfield.

4.13 Hilbert-Ackermann's Theorem (IS_1). Let T be an open theory (all axioms open). If T is inconsistent then there is a disjunction D of instances of negations of axioms of T such that D is a propositional tautology.

Remark. An instance of a formula $\varphi(x_1 \dots x_n)$ is any formula $\varphi(t_1, \dots, t_n)$ where $t_1, \dots, t_i$ are terms, t_i free for x_i in φ . An open formula is a propositional tautology if each evaluation of its atoms by zeros and ones produces the value 1 using the truth tables of connectives.

Checking the fact that Shoenfield's proof formalizes in IS_1 is tedious but straightforward and we shall omit it. Note that we assume the equality axioms to be particular axioms of T so that our tautologies are just Shoenfield's quasi-tautologies.

4.14 Corollary (IS_1). A closed existential formula is provable (in predicate calculus) iff there is a disjunction of instances of its open part which is a propositional tautology.

4.15 Herbrand's Theorem (IS_1). A formula φ is provable in predicate calculus iff its Herbrand normal form is provable in predicate calculus iff there is a disjunction of closed instances of the open part of $He(\varphi)$ that is a propositional tautology.

Remark. Here again we refer to the proof in Shoenfield's book; the reader may check without difficulty that this proof formalizes in IS_1 . See 0.21 and

III.3.30. Note again that Herbrand's theorem is also analyzed in Chap. V where it is shown that it is provable a theory weaker than $I\Sigma_1$.

4.16 Theorem ($I\Sigma_1$). If T is a theory in L , $T \vdash (\forall x)(\exists y)\varphi(x, y)$ and F is a new function symbol then $T \cup \{(\forall x)\varphi(x, F(x))\}$ extends T conservatively.

Proof. As in [Shoenfield]. Take F to be $F_{\varphi, y}$ and observe that it suffices to show that whenever $(\forall x)\varphi(F(x)) \rightarrow \alpha$ is provable in predicate calculus (where α is a L -formula) then $(\forall x)(\exists y)\varphi(x, y) \rightarrow \alpha$ is also provable. To see this it is enough to observe that both formulas have the same Herbrand normal form. $\square$

4.17 Definition ($I\Sigma_1$). $Sk(T) = \{Sk(\varphi) \mid \varphi \in T\}$.

4.18 Corollary ($I\Sigma_1$). $Sk(T)$ is a conservative extension of T .

Proof. Routine, use 4.16 (similarly to the proof of 4.10, i.e. using $L\Sigma_1$). $\square$

4.19 Corollary ($I\Sigma_1$). T is consistent iff each finite set of closed instances of $Sk(T)$ is propositionally satisfiable.

Proof. T is consistent iff each finite $t \subseteq T$ is consistent. Such a $t = \{\varphi_0, \dots, \varphi_v\}$ is consistent iff $\{Sk(\varphi_0), \dots, Sk(\varphi_v)\}$ is consistent (by 4.18); this is equivalent to the propositional satisfiability of each finite set of closed instances of $Sk(\varphi_0), \dots, Sk(\varphi_v)$ by Hilbert-Ackermann. $\square$

4.20 Theorem ($I\Sigma_1$). Each theory T has a conservative open extension T' in which each formula is equivalent to an open formula. If T is Δ_1 then T' is also Δ_1 .

Proof. Like 0.17, using 4.11 (2) (construction of the language). $\square$

(b) Arithmetizing Model Theory

Here we shall define models for a language and prove some theorems on the relation between consistency and existence of a model. The notion of a partial satisfaction plays a prominent role.

4.21 Definition ($I\Sigma_1$). Let L be a language. A *model* for L consists of a non-empty set M and a system $S = ((R_P)_{P \in Pred}, (f_F)_{F \in Fct}, (m_c)_{c \in const})$ such that for each $P \in Pred$, $R_P \subseteq M^x$ where x is the arity of P , $f_F: M^x \rightarrow M$ where x is the arity of F and m_c is an element of M . As usual, we write M instead of (M, S) .

Remark. This sounds very familiar but the reader should keep in mind two things:

(1) In fact this is a scheme of definitions depending on the range of the variable M . We may speak on Δ_1 models, Σ_{17} models, etc; working in $B\Sigma_2$, we may speak on low Δ_2 models etc.

(2) S is a system of sets indexed by elements of a set; this is easily expressed by requiring that S be a set of ordered pairs and, for example, if $P \in \text{Pred}$, then $R_P = \{x \mid (x, P) \in S\}$: saying that R_P is a z -ary relation we mean that R_P consists of sequences s of elements of M such that $lh(s) = z$. A mapping $f : M^x \rightarrow M$ is understood to be a particular $(x+1)$ -ary relation on M .

(Definition continued). Let M be a model for a language L . Define an *evaluation of variables of a term* by copying Def. 0.4 in $I\Sigma_1$ (and generalizing 1.64): e is an evaluation of variables of G in M if e is a finite mapping associating to some variables (among them all variables of t) elements of M . A function Val is an *evaluation of terms in M* if it satisfies the conditions of 0.4 (copied in $I\Sigma_1$), i.e. $Val(u, e) = e(u)$ if u is a variable, $Val(u, e) = u_M$ if u is a constant, $Val(F(t_1 \dots t_x), e) = f_F(Val(t_1, e), \dots, Val(t_x, e))$.

4.22 Lemma ($I\Sigma_1$). If M is Δ_1 then there is a unique Δ_1 evaluation of terms in M .

Proof. Routine. (Note that we assume M to be Δ_1 .) □

4.23 Definition ($I\Sigma_1$). Let M be a model for L and let Val be a valuation of terms. Let Λ be a set of L -formulas closed under the taking of subformulas and substituting terms. A relation Sat is a *partial satisfaction* for Λ in M if it satisfies Tarski's conditions w.r.t. M and Λ (copy 0.4). Sat is a *full satisfaction* for M if it is a satisfaction for all L -formulas in M . (M, Sat) is a *full model* if M is a model and Sat is a full satisfaction in M .

Under the notation of 4.23, a Λ -formula is *true* in (M, Sat) if for each evaluation e of variables of z , $(z, e) \in Sat$. If T is a theory and each axiom of T is a Λ -formula then M is a *model* of T if each axiom of T is true in (M, Sat) .

4.24 Theorem ($I\Sigma_1$). If M is a model for L and M is Δ_1 then there is a unique $S \in \Delta_1$ which is a satisfaction for open formulas in M .

Proof. Routine (imitate 1.71–1.74; you have to relate everything to M but you do not have to deal with quantifiers). □

4.25 Theorem ($I\Sigma_1$). If T is a decidable theory (i.e. the set of all provable formulas is Δ_1) then T has a full Δ_1 model.

Proof. Just imitate the usual proof of the completeness theorem: first extend T conservatively to a Henkin theory $T' \in \Delta$ (recall that T' is decidable) and then extend T' to a complete decidable theory T'' having the same language as T' . To this end, use a Δ_1 increasing enumeration Z of closed T' -formulas (write z_i for $Z(i)$ and define $H(i) = z_i$ if $\text{Con}^*(T' \cup \{H(0), \dots, H(i-1), z_i\})$, $H(i) = \neg z_i$ otherwise; put $T'' = \text{range}(H)$, i.e. $z_i \in T''$ iff $T'' \vdash z_i$ iff $H(i) = z_i$. Thus T'' is decidable. For each closed term t let $E(t)$ be the least closed term t' such that $T'' \vdash t =^* t'$; let $M = \text{range}(E)$. Observe that M is Δ_1 ($t \in M$ iff $T'' \vdash E(t) =^* t$); for $t_1, \dots, t_x \in M$ put $(t_1 \dots t_x) \in R_P$ iff $T'' \vdash^* P(t_1 \dots t_x)$ and put $f_P(t_1 \dots t_x) = E(F(t_1 \dots t_x))$. If z is a formula and $u_0 \dots u_x$ are its free variables then put

$$\text{Sat}(z, [t_0 \dots t_x]) \equiv T' \vdash^* \text{Subst}^*(z, (u_0 \dots u_x), (t_1 \dots t_x)))$$

(x -fold substitution). Then show by induction on formulas that Tarski's truth conditions are obeyed, observing that they are $\Sigma_0(\Sigma_1)$ (condition for $\forall$!); thus $I\Sigma_1$ is sufficient. $\square$

4.26 Theorem ($I\Sigma_1$). If M is a Δ_1 model and $T = \text{Tr}_{\text{open}}(M)$ is the set of all open formulas true in M then $\text{Con}^*(T)$.

Proof. By Hilbert-Ackermann's theorem, if T is inconsistent then there is a disjunction δ of instances of negations of axioms of T that is a propositional tautology. Let Sat be the satisfaction for open formulas in M ; prove by Δ_1 -induction that δ is true in M . (For each evaluation of variables of δ compute the corresponding evaluation of atoms of δ by zeros and ones; 1 means that the atom is satisfied). Show that for each open formula built from these atoms we get: φ is satisfied by our evaluation in M iff our propositional evaluation gives φ the value 1.)

On the other hand, each instance of our axiom is true in M , and the same holds for any conjunction of such instances (again Σ_1 induction suffices). Thus $\neg\delta$ is true in M and also δ ; this is a contradiction. Thus T is consistent. $\square$

4.27 Low Arithmetized Completeness Theorem.

- (1) ($I\Sigma_1$) If $T \in \Delta_1$ is a theory then T is consistent iff it has a full low $\Sigma_0^*(\Sigma_1)$ model.
- (2) ($B\Sigma_2$) If $T \in \Delta_1$ is a theory then T is consistent iff it has a full low Δ_2 model.

Proof. Let $T \in \Delta_1$; we may assume T to be Henkin. Define a dyadic tree $\text{Tree}(T)$ as follows: first take a Δ_1 enumeration of all sentences of L ; denote it by $(z_0, z_1, \dots)$. Put $(0)z = \neg^* z$ and $(1)z = z$. For each string s , let $s \in \text{Tree}(T)$ iff there is no $p \leq s$ such that p is a T -proof of a contradiction from $\{((s)_i)z_i \mid i < \text{lh}(s)\}$.

Thus s is in the tree iff there is no short proof of a contradiction from the first $lh(s)$ sentences negated or asserted according to s . We show by Σ_1 -induction that the tree is unbounded, namely we show that for each x there is an s such that $Con(\{(s)_i z_i \mid i < x\})$. (The induction step is evident.) Thus the Low Basis Theorem gives an infinite branch $B = (\varepsilon_0, \varepsilon_1, \dots)$.

The branch is LL_1 (i.e. $\Sigma_0^*(\Sigma_1)$). Use the Δ_1 enumeration of all closed formulas above: we define $\hat{T} = \{z_n \mid \varepsilon_n = 1\}$. Clearly, $\hat{T}$ is $\Delta_1(B)$ and hence LL_1 . $\hat{T}$ defines the canonical structure as follows: $M = \{c \text{ constant} \mid (\forall c' < c, c' \text{ constant})(c \neq^* c') \in \hat{T}\}$ (representatives of classes of provably equal constants) – note that M is $\Delta_1(\hat{T})$ and hence LL_1 . Predicates and function symbols are interpreted as follows:

$$(c_0, \dots, c_x) \in R_P \text{ iff } P(c_0, \dots, c_x) \in \hat{T};$$

$$d = f_F(c_0, \dots, c_x) \text{ iff } (d =^* F(c_0, \dots, c_x)) \in \hat{T}.$$

Clearly the structure M is LL_1 ; one defines $Sat_M(z, e)$ iff $Subst^*(z, e) \in \hat{T}$ (we substitute the constants assigned by e for the free variables of z). Clearly, Sat_M is LL_1 and one verifies easily Tarski truth conditions for connectives and quantifiers. Thus (M, Sat_M) is a full LL_1 model, and for each (closed) axiom z of T , $Sat_M(z, \emptyset)$; thus (M, Sat_M) is a model of T .

Conversely, assume that (M, Sat) is a full LL_1 model of T . Let $z_0, \dots, z_x$ be a T -proof; show as usual by induction that each member of this sequence is true in (M, Sat) . The induction step is routine; and the statement is $(\forall e \text{ evaluation of } z_i)(Sat(z_i, e))$, thus $\Pi_1(Sat)$, *a fortiori*, $\Sigma_0^*(\Sigma_1)$ and the result follows by $I\Sigma_0^*(\Sigma_1)$. (Alternatively, we can prove that for each i , the universal closure $(\forall^* \dots)z_i$ of z_i is satisfied by $\emptyset$, which is $\Delta_1(Sat)$.) $\square$

4.28 Problem. Can we formalize the trivial model-theoretic proof of 0.14? This seems to be a peculiar task; starting from a LL_1 model you can interpret a Skolem function by an LL_1 mapping as in 0.14; but does the expanded structure have an LL_1 full satisfaction?

(c) Applications to Arithmetic

In subsections (a), (b) we arithmetized some parts of logic of axiomatized theories; in (a) we dealt with elements of proof theory and in (b) with elements of model theory. This logic has been developed in $I\Sigma_1$. In this subsection we apply this framework to particular arithmetized theories, namely to fragments of arithmetic themselves. We shall introduce Δ_1 theories $I\Sigma_k^*$, $B\Sigma_k^*$, PA^* etc. and derive some of their properties in $I\Sigma_1$. We show that, for each k , $I\Sigma_{k+1} \vdash Con^*(I\Sigma_k^*)$ (theorem 4.34). Formalization of parts of the model theory of fragments presupposes satisfactory knowledge of the non-formalized (actual) model theory of fragments, which will be studied in Chap. IV. Thus

we postpone our discussion of important parts of arithmetized metamathematics of fragments until that chapter. Note that by using formalized model theory we shall be able to strengthen the above result and prove that, for each k , $I\Sigma_{k+1} \vdash \text{Con}^*(B\Sigma_{k+1}^*)$ (Theorem IV.4.8). And Gödel's second incompleteness theorem (III.2.21) gives the unprovability of $\text{Con}(I\Sigma_{k+1}^*)$ in $I\Sigma_{k+1}$.

4.29 Definition ($I\Sigma_1$). Define in the obvious way Δ_1 functions assigning to each formula z of the language of arithmetic and to each variable u the corresponding scheme I^*z and collection scheme B^*z ; define the finite set Q^* of axioms of Robinson's arithmetic as $\{\bar{Q}_0, \dots, \bar{Q}_8\}$, where Q_i are axioms of Q . Define $I\Sigma_x^*$ to be the set $Q^* \cup \{I^*z \mid z \in \Sigma_x^*\}$ and similarly for $B\Sigma_x^*$, PA^* , etc.

4.30 Remark. For each k , the set $I\Sigma_k^*$ is provably Δ_1 and obviously defines $I\Sigma_k$ in N ; evidently, the formula $(z \in I\Sigma_k^*)$ binumerates $I\Sigma_k$ in $I\Sigma_1$, i.e. for each formula α , α is an axiom of $I\Sigma_k$ iff $I\Sigma_1 \vdash \alpha \in I\Sigma_k^*$, and α is not an axiom of $I\Sigma_k$ iff $I\Sigma_1 \vdash \alpha \notin I\Sigma_k^*$.

Binumerations of (actual) axiom systems will be studied closely in Chap. III; here we have for each of our systems just one particular binumeration, whose definition in $I\Sigma_1$ only copies the actual definition we gave in Sect. 1 or 2.

4.31 Remark. We have a particular model for the language of arithmetic, namely the *universe* (the Δ_1 set V of all numbers) endowed with the Δ_1 relation Ord' defined by $x \leq y$, the Δ_1 functions succ ($\text{succ}(x) = S(x)$), add ($\text{add}(x, y) = x + y$) and mult ($\text{mult}(x, y) = x * y$) and the element 0. We can also speak of the *standard model* but the reader should keep in mind that everything is said and meant *inside* of $I\Sigma_1$. Recall satisfactions $\text{Sat}_{\Sigma, k}$ and $\text{Sat}_{\Pi, k}$: they are partial satisfactions for the universe in the sense of 4.22.

4.32 Theorem ($I\Sigma_1$ -formalized Σ_1 -completeness). Q^* proves each true Σ_1^* -sentence, i.e. if $z \in \Sigma_1^*$ is closed and $\text{Sat}_{\Sigma, 1}(z, \emptyset)$ then $Q^* \vdash z$.

Proof. Copy the proof of 1.8; it suffices to prove the assertion for $z \in \Sigma_0$. Since $Q^* \vdash z$ is a Σ_1 -formula the proof by induction formalizes as it stands. $\square$

4.33 Theorem. For each $k \geq 1$, $I\Sigma_k$ proves the consistency of the set of all true Π_{k+1}^* sentences, i.e. if $\text{Tr}(\Pi_{k+1}^*)$ is the Π_{k+1} -set of all true Π_{k+1}^* -sentences then $I\Sigma_k \vdash \text{Con}^*(\text{Tr}(\Pi_{k+1}^*))$. (More pedantically, $\text{Tr}(\Pi_{k+1}^*)$ is a formula with the only free variable x saying “ x is a closed formula” and $\text{Sat}_{\Pi, k+1}^*(x, \emptyset)$.”.)

Proof. The idea of the proof is rather simple: in $I\Sigma_k$, take a finite set π of true $\Pi_{k+1}^\bullet$ formulas and expand the standard model to a Δ_k model of $Sk(\pi)$. Since $Sk(\pi)$ is open, we get $Con^\bullet(Sk(\pi))$ by relativizing 4.26 to $I\Sigma_1(\Delta_1) = I\Sigma_1$. And we get $Con^\bullet(\pi)$ since $Sk(\pi)$ proves π .

We shall elaborate on this. Let π be our finite set of true $\Pi_{k+1}^\bullet$ -formulas and let Φ be one of them. We may assume that the axioms of linear order appear among them; then we may assume without loss of generality that each Φ has its bounded part in bounded prenex form, i.e. Φ consists of a block of unbounded alternating quantifiers (the first being $\forall$) followed by a block of bounded quantifiers followed by an open formula.

Follow the definition of the Skolem normal form in 4.11. Our Φ may be presented as follows:

$$(\forall^\bullet u_1)(\exists^\bullet u_2) \dots (Q_{k+1} u_{k+1})(Q_{k+2} u_{k+2} \leq^\bullet v_{k+2}) \dots (Q_r u_r \leq^\bullet v_r) \varphi(u, v)$$

where each v_i is some u_j , $j < i$.

For $i = 0, \dots, p$, let $\Phi^{(i)}$ be the result of deleting the i leftmost quantifiers (binding $u_1, \dots, u_i$) in Φ . Recall the terms $t_1, \dots, t_p$ in 4.11: if Q_j is $\exists^\bullet$ then t_j is $F_j^\Phi(\leftarrow t_{j-1})$ and if Q_j is a $\forall^\bullet$ then t_j is u_j . We define

$$f_1^\Phi(x_1) = (\text{least } x_2) \text{Sat}_{\pi, k-1}(\Phi^{(2)}, [x_1, x_2]);$$

note that $\Phi^{(2)}$ is just $(Q_3 u_3 \dots) \varphi(u)$, which is Σ_{k-1} ; thus f_1^Φ is total since Φ is true and by $L\Sigma_{k-1}$. Furthermore, f_1^Φ is Δ_k . For all other i such that Q_{i+1} is $\exists^\bullet$ we define

(a) for $i \leq k$:

$f_i^\Phi(\leftarrow x_i) = (\text{least } x_{i+1}) \text{Sat}_{\Pi, k-1}(\Phi^{(i+1)}, [x_1, \dots, x_{i+1}])$ if there is such an x_{i+1} , and = 0 otherwise.

(b) for $i > k$:

$f_i^\Phi(\leftarrow x_i) = (\text{least } x_{i+1} \leq x_j) \text{Sat}_0^{\Phi^{(i+1)}}[x_1 \dots x_{i+1}]$ if there is such an x_{i+1} , = 0 otherwise (here v_{i+1} is u_j).

Observe that $i \geq 3$ so that f_i^Φ is certainly Δ_k . We consider the Δ_k extension of the standard model by all the (finitely many) f_i^Φ as interpretations of F_i^Φ . Recall that we have a Δ_k valuation Val^* of terms and a Δ_k satisfaction Sat^* for open formulas of the extended language (relativized 4.24). $\square$

Claim. Let $[x]$ be an evaluation of variables^{*} $u_1, \dots, u_q$ by numbers $x_1, \dots, x_q$ such that for each $i = 1, \dots, k$ we have $x_i = Val^*(t_i, [x])$; then $Sat_0(\varphi(u_1 \dots u_k), [x])$.

Prove the claim by showing the following by induction on $j = 0, \dots, q$: $Sat_{\Pi, k-1}(\Phi^{(j)}, [x])$. We have induction enough at our disposal; and the induction step just uses the definition of F_i^Φ .

Note that the evaluation $[x]$ satisfying the condition of the claim is fully determined by its members x_i such that Q_i is $\forall^\bullet$; let $[x]'$ be the correspond-

ing subevaluation. The claim gives evidently the following: for each evaluation $[x]'$, $Sat^*(\varphi(t_1, \dots, t_q), [x]')$. (This is because we get $Val^*(t_i, [x]') = Val^*(t_i, [x]) = x_i$; recall that φ is open).

Consequently, we have shown that for each Φ in π and each evaluation e of free variables of $Sk(\pi)$, e satisfies $Sk(\pi)$ in Sat^* , hence $Sk(\pi)$ is consistent and π is also consistent. (Once again: consistency of $Sk(\pi)$ uses Hilbert-Ackermann, see 4.26; consistency of π follows since π is provable from $Sk(\pi)$ in logic.)

4.34 Corollary. (1) For each k , $I\Sigma_{k+1} \vdash Con(I\Sigma_k^*)$. (This is because $I\Sigma_k^*$ is Π_{k+2}^* -axiomatized.)

(2) Moreover, $I\Sigma_{k+1}$ proves the following: if z is a true Σ_{k+3}^* -sentence (i.e. $Sat_{\Sigma, k+3}(z, \emptyset)$) then $Con^*(I\Sigma_k \cup \{z\})$. (This is because a true Σ_{k+3}^* sentence z is implied by a true Π_{k+2}^* -sentence in which the leftmost existential quantifier of z is replaced by a witness.)

(3) (Reflection). For each Σ_{k+3} formula φ , $I\Sigma_{k+1} \vdash \varphi \rightarrow Con^*(\varphi)$. (Thus for each φ , $PA \vdash \varphi \rightarrow Con(\varphi)$.)

4.35 Remark. (1) This cannot be improved on by allowing φ to be also Π_{k+3} since $I\Sigma_{k+1}$ is itself axiomatized by a single Π_{k+3} sentence and we would have $I\Sigma_{k+3} \vdash Con^*(I\Sigma_{k+3})$ which contradicts Gödel's second incompleteness theorem.

(2) Note that in Chap. IV we prove $I\Sigma_{k+1} \vdash Con^*(B\Sigma_{k+1}^*)$; this will use the formalized model theory of fragments.

(3) In the rest of this section, we prove a theorem which is a modification of 4.33 (or at least its proof is a modification of the proof of 4.33). It is a sufficient condition for consistency using Δ_1 models. We shall use it in Chap. II in connection with Paris-Harrington principle.

4.36 Definition. ($I\Sigma_1$). The skolemization of a formula Φ in the language of arithmetic with respect to unbounded quantifiers is defined similarly to the full skolemization, but bounded quantifiers are left untouched. Thus: let Φ be $(Q_1 x_1) \dots (Q_k x_k) \varphi(\mathbf{x}, \mathbf{y})$ where φ is Σ_0^* and Q_i is $\forall^*$ or $\exists^*$. Let

$$\begin{aligned} t_i &= x_i \text{ if } Q_i \text{ is } \forall^*, \\ t_i &= F_i^\Phi(\leftarrow t_{i-1}) \text{ if } Q_i \text{ is } \exists^*. \end{aligned}$$

Then $Sk_0(\Phi)$ is $\varphi(t_1 \dots t_k)$. A closed instance of $Sk_0(\Phi)$ is any formula $\varphi(s_1, \dots, s_u, \mathbf{u})$ where $\mathbf{s}, \mathbf{u}$, are tuples of closed terms in the extended language, and if Q_i is $\exists^*$ then s_i is $F_i^\Phi(\leftarrow s_{i-1}, \mathbf{u})$. For each set T of formulas, $Sk_0(T) = \{Sk_0(\Phi) \mid \Phi \in T\}$.

4.37 Theorem ($I\Sigma_1$). Let T be a Δ_1 theory in the language of arithmetic. Assume that for each finite set S_0 of closed instances of $Sk_0(T)$ there is a Δ_1 expansion of the standard model to a model of S_0 . Then $Con(T)$.

Proof. First two remarks. (1) Recall 4.13 implying that if each finite set of closed instances of $Sk(T)$ (full skolemization) is true in a Δ_1 expansion of the standard model then $Con(T)$ (since truth of an open-closed formula in a Δ_1 model gives trivially propositional satisfaction).

(2) We know that a Δ_1 structure has a Δ_1 satisfaction for open formulas; observe that if we Δ_1 expand the standard model by adding interpretation of (finitely many) functions then the satisfaction Sat_0 immediately extends to a Δ_1 satisfaction for formulas resulting from $\Sigma_0^\bullet$ formulas by substituting $L^\bullet$ -terms for their variables ($Sk_0(\emptyset)$ has such a form): define for $[x] = [x_1 \dots x_k]$

$$Sat_0(\varphi(t_1, \dots, t_k), [x_1 \dots x_k]) \text{ iff} \\ Sat_0(\varphi(u_1 \dots u_k), [Val(t_1[x]), \dots, Val(t_k, [x])]).$$

Trivially, Tarski conditions are satisfied. Saying that a Δ_1 structure expanding the standard model is a model of S_0 means that each element of S_0 is true in the extension of Sat_0 just described.

The proof of 4.37 may now follow the same pattern as that of 4.33 with the only difference being that the Skolem functions for unbounded quantifiers need not be constructed since they are given (as Δ_1 functions) in advance:

Take a finite set S of closed instances of $Sk(T)$ and find a finite set S_0 of closed instances of $Sk_0(T)$ such that S is a set of instances of $Sk(S_0)$. Let M be a Δ_1 expansion of the standard model in which S_0 is true. Now define Δ_1 interpretations of Skolem functions of $Sk(S_0)$ (we deal with bounded quantifiers now – cf. 4.33 for $i > k$) and get a Δ_1 expansion M' of M in which S is true. $\square$

4.38 Remark. The same holds for the case where T is formulated in the language with predicates for the successor, addition and multiplication, cf. Sect. 2 (e).

Chapter II

Fragments and Combinatorics

Introduction. In the present chapter we shall elaborate proofs of various combinatorial principles in suitable fragments of arithmetic. In general, infinite principles deal with graphs, functions etc. on infinite sets, finite principles relate similarly to finite sets. We prove both some infinite and some finite principles; furthermore, we show some infinite principles to be equivalent to certain collection principles and some finite principles to be equivalent to certain consistency statements. Sections 1 and 2 deal with strengthenings of the infinite and finite Ramsey theorem (they will be formulated at the beginning of Sect. 1), in particular with various forms and instances of Paris-Harrington principle. This principle is very famous since it has been the first example of an arithmetical statement that has a clear combinatorial meaning, is true (in N) and is unprovable in PA .

Instances of Paris-Harrington principle will form a hierarchy of formulas, n -th of them will be proved in $I\Sigma_n$ ($n \geq 1$). As said above, in this chapter we deal with concrete proofs, not with unprovability; but unprovability results immediately follow from the results of this chapter using Gödel's incompleteness theorems (elaborated in Chap. III). We shall mention this on corresponding places in this chapter: $(n+1)$ -th instance is unprovable in $I\Sigma_n$.

In Sect. 3 we shall deal with ordinals in fragments, introduce the notion of α -large sets (α an ordinal) and investigate another hierarchy of combinatorial statements, related to the first one. Results of this section will be used in Chap. IV for a characterization of functions provably recursive in $I\Sigma_n$ ($n \geq 1$).

1. Ramsey's Theorems and Fragments

(a) Statement of Results

1.1 First we shall recall Ramsey's theorems in an informal formulation. If X is a set of natural numbers, then $[X]^u$ is the set of all u -element subsets

of X (or, equivalently, all increasing u -element sequences of elements of X). $F : [X]^u \rightarrow a$ (where a is a natural number) means that F is a mapping whose domain is $[X]^u$ and whose range is included in $\{0, 1, \dots, a-1\}$. It is customary to call u the *arity* of F and a the *number of colours*. $Y \subseteq X$ is *homogeneous* for F if F restricted to $[Y]^u$ is constant. The *infinite Ramsey theorem* says that for each natural u and a and each $F : [X]^u \rightarrow a$ where X is unbounded (i.e. infinite) there is an unbounded $Y \subseteq X$ which is homogeneous for F . It is customary to denote this by

$$(\forall u, a)(\omega \rightarrow (\omega)_a^u) \quad \text{or} \quad (\forall u)(\omega \rightarrow (\omega)_{<\omega}^u);$$

the first ω symbolizing the unboundedness of X and the second unboundedness of the homogeneous set.

1.2 For natural numbers, x, y, u, z, q , the symbol $[x, y] \rightarrow (q)_z^u$ means that if X is the closed interval $[x, y]$ of natural numbers, then each $f : [X]^u \rightarrow z$ has a homogeneous set of cardinality q . The *finite Ramsey theorem* is

$$(\forall x, q, u, z)(\exists y)([x, y] \rightarrow (q)_z^u).$$

The symbol $[x, y] \rightarrow_* (q)_z^u$ means that each f as above has a homogeneous set Y of cardinality q which is *relatively large*, i.e. $\min(Y) \leq \text{card}(Y)$. *Paris-Harrington principle* is

$$(\forall x, q, u, z)(\exists y)([x, y] \rightarrow_* (q)_z^u).$$

Evidently, this is a strengthening of the finite Ramsey theorem; Paris-Harrington principle follows from the infinite Ramsey theorem using König's lemma. (We shall discuss the proof below.)

1.3 Clearly, Paris-Harrington principle as well as the finite Ramsey theorem is expressible by a formula of first order arithmetic; let us write $PH(u, z)$ for $(\forall x, q)(\exists y)([x, y] \rightarrow_* (q)_z^u)$ (recall that u is the arity and z the number of colours). Thus Paris-Harrington principle is $(\forall u, z)PH(u, z)$. Our sequence of formulas that are "harder and harder to prove" is $(\forall z)PH(\bar{n}, z)$ for $n = 1, 2, \dots$

1.4 On the other hand, we cannot formalize the infinite Ramsey theorem in first-order arithmetic as it stands since we cannot quantify over arbitrary sets of natural numbers. But we can quantify over sets of restricted complexity, e.g. over Δ_m sets (in IS_1) or *low* Δ_m (in $B\Sigma_m$). Thus we may express several partition relations saying that for each Γ_1 -definable and unbounded X and for each Γ_1 -definable $F : [x]^u \rightarrow z$ there is a Γ_2 -definable unbounded

homogeneous set (where Γ_1, Γ_2 are Δ_m , low Δ_m or so). Denote such a formula by

$$\omega \rightarrow (\omega)_z^u(\Gamma_1, \Gamma_2).$$

Recursive analysis of Ramsey's theorem consists in establishing truth of assertions of this type. (Pioneering work was done by Jockusch). Our aim is still more ambitious: we want to establish provability of such assertions in suitable fragments of arithmetic. We are now ready to present the main results of this section.

1.5 Theorem. For $m, n \geq 1$, $B\Sigma_{m+n}$ proves

$$\omega \rightarrow (\omega)_{<\omega}^n(\text{low } \Delta_{m+1}, \text{low } \Delta_{m+1}),$$

i.e.: For each z , if X is low Δ_{m+1} and unbounded and F is a low Δ_{m+1} mapping of $[X]^n$ into z (i.e. into $\{0, \dots, z-1\}$) then F has a low Δ_{m+1} homogeneous unbounded set.

For a proof (using low basis theorem) see below. Note that the assertion is *meaningful* in $B\Sigma_{m+1}$ and is expressible as a single formula using the coding of low Δ_{m+1} sets in $B\Sigma_{m+1}$ (see Chap. I, Sect. 2). Due to some obvious inclusions, we have e.g. the following corollary: for $m, n \geq 1$, $B\Sigma_{m+n}$ proves

$$\omega \rightarrow (\omega)_{<\omega}^n(\Delta_m, \Delta_{m+n}).$$

This assertion is weaker but is meaningful already in $I\Sigma_1$ and is equivalent over $I\Sigma_1$ to $B\Sigma_{m+n}$.

1.6 Theorem. For $m, n \geq 1$, $I\Sigma_1$ proves the following:

$$B\Sigma_{m+n} \equiv \omega \rightarrow (\omega)_{<\omega}^n(\Delta_m, \Delta_{m+1}).$$

(Here $B\Sigma_{m+n}$ is formulated as a single formula).

By Theorem 1.5, $B\Sigma_{m+n}$ proves an infinite Ramsey type theorem on mappings of arity n and complexity low Δ_{m+1} . We shall see that this theory also proves a Ramsey type theorem on mappings of arity $(n+1)$ and complexity low Δ_{m+1} ; but it guarantees only a finite relatively large homogeneous set and the number of colours must be standard.

1.7 Theorem. Let $m, n \geq 1$. (1) $I\Sigma_{m+n-1}$ proves the following: If X is LL_m and unbounded and if $F : [X]^n \rightarrow z$ is LL_m , then for each q there is a relatively large homogeneous (finite) set having at least q elements. This can be expressed by

$$(\forall z)(\forall q)(\omega \rightarrow (q)_z^n(LL_m)).$$

(2) For each k , $I\Sigma_{m+n-1}$ proves

$$(\forall q)(\omega \rightarrow_*(q)_k^{n+1}(LL_m)).$$

$I\Sigma_{m+n}$ also proves that infinite homogeneous sets of some complexity need not exist. As an example we prove the following.

1.8 Theorem. For $m \geq 1$, $I\Sigma_{m+1}$ proves

$$\neg\omega \rightarrow (\omega)_2^2(\Delta_m, \Sigma_{m+1})$$

(Thus there is a Δ_m $F : [X]^2 \rightarrow 2$, $X \Delta_m$, unbounded, with no Σ_{m+1} homogeneous unbounded set).

Note that a stronger result will be obtained in 1.28.

Let us summarize the above results for arity 2, 2 colours and Δ_1 mappings. We have the following:

$$\begin{aligned} I\Sigma_1 &\vdash (\forall q)\omega \rightarrow_*(q)_2^2(\Delta_1), \\ I\Sigma_2 &\vdash \neg\omega \rightarrow (\omega)_2^2(\Delta_1, \Sigma_2), \\ B\Sigma_3 &\vdash \omega \rightarrow (\omega)_2^2(\Delta_1, \text{low } \Delta_3). \end{aligned}$$

Now let us consider finite Ramsey type theorems.

1.9 Theorem. For $n \geq 1$,

- (a) $I\Sigma_n \vdash (\forall x, z, q)(\exists y)([x, y] \rightarrow_*(q)_z^{\bar{n}})$
(i.e. $I\Sigma_n \vdash (\forall z)PH(\bar{n}, z)$),
- (b) for each k ,
 $I\Sigma_n \vdash (\forall x, q)(\exists y)([x, y] \rightarrow_*(q)^{\bar{n}+1})$
(i.e. for each k , $I\Sigma_n \vdash PH(\bar{n}+1, \bar{k})$).

Remark. Results of Chaps. III and IV enable us to add the following:

First, all formulas in question are Π_2 .

Second, in Sect. 2 of the present chapter we prove a theorem implying that $I\Sigma_1$ proves $(\forall z)PH(\bar{n}+1, z) \rightarrow \text{Con}(I\Sigma_n^*)$ (consistency), thus we may apply Gödel's second incompleteness theorem (proved in III.2.21) to deduce the unprovability of $(\forall z)PH(\bar{n}+1, z)$ in $I\Sigma_n$. Thus we have a strictly increasing hierarchy of Π_2 formulas.

1.10 Theorem. $I\Sigma_1$ proves $(\forall x, z, q, u)(\exists y)([x, y] \rightarrow (q)_z^u)$.

This completes our list of results. In what follows we shall elaborate proofs.

(b) Proofs (of 1.5, 1.7, 1.9)

1.11 Definitions (IS_1). Let $F : [X]^2 \rightarrow a$ be Δ_1 , X Δ_1 unbounded. An increasing sequence s of elements of X is *prehomogeneous* if for each $i < j < lh(s)$ the value $F((s)_i, (s)_j)$ depends only on the first argument, i.e. for each $i < j < k < lh(s)$ we have $F((s)_i, (s)_j) = F((s)_i, (s)_k)$. If $i < lh(s) - 1$ then the *colour* of $(s)_i$ (in s) is the common value $F((s)_i, (s)_j)$ for $i < j < lh(s)$. If $lh(s) > 0$, $u > \max(s)$ and $s \smallfrown \langle u \rangle$ is prehomogeneous, then the maximal element of s has a colour in $s \smallfrown \langle u \rangle$, namely $F(\max(s), u)$. Let $lh(s) \geq 1$; $s \smallfrown \langle u \rangle$ is a *minimal prehomogeneous extension* of s if $s \smallfrown \langle u \rangle$ is prehomogeneous and there is no v between $\max(s)$ and u such that $s \smallfrown \langle v \rangle$ is prehomogeneous and $\max(s)$ has the same colour in $s \smallfrown \langle u \rangle$ as in $s \smallfrown \langle v \rangle$. s is *hereditarily minimal prehomogeneous* (or h.m.p.h.) if s is prehomogeneous, $s = \emptyset$ or $(s)_0 = \min X$ and for each i between 1 and $lh(s) - 1$, $s \upharpoonright (i + 1)$ is a minimal prehomogeneous extension of $s \upharpoonright i$ is the initial segment of s of length i).

1.12 Definition. (IS_1). A tree T is *narrowly branching* if there is a number c such that each $s \in T$ has at most c immediate successors.

1.13 Lemma. (IS_1) Let $F : [X]^2 \rightarrow a$ be as in 1.11 and let T be the set of all h.m.p.h. sequences. Then T is Δ_1 , and is an unbounded narrowly branching tree.

Proof. Evidently, T is a Δ_1 tree. It is narrowly branching since each $s \in T$ has at most a immediate successors. To prove that T is unbounded one easily shows that for each $x \in X$ there is an $s \in T$ such that $\max(s) = x$. Indeed, let $s_0 = \min X$; then $\langle s_0, x \rangle$ is trivially prehomogeneous and $\langle s_0 \rangle$ is h.m.p.h. Assume we have a h.m.p.h. sequence s such that $s \smallfrown \langle x \rangle$ is prehomogeneous. If $s \smallfrown \langle x \rangle$ is not h.m.p.h., then there is a $y < x$ such that $s \smallfrown \langle y \rangle$ is h.m.p.h. and the colour of $\max(s)$ in $s \smallfrown \langle y \rangle$ is the same as the colour of $\max(s)$ in $s \smallfrown \langle x \rangle$; then $s \smallfrown \langle y \rangle \smallfrown \langle x \rangle$ is prehomogeneous. $\square$

1.14 Lemma (1) (IS_1). If T is a Δ_1 narrowly branching unbounded tree, then for each level x , the set $\{s \in T \mid lh(s) = x\}$ is bounded.

(2) ($B\Sigma_2$). If T is as above then for each level x there is an $s \in T$ such that $lh(s) = x$ and the Δ_1 set $T_s = \{t \in T \mid t \supseteq s\}$ is unbounded. Thus $\{t \mid s \smallfrown t \in T\}$ is an unbounded Δ_1 narrowly branching tree.

Proof. (1) Associate with each $s \in T$ a sequence $H(s) = s'$ of the same length defined as follows: $\emptyset'$ is $\emptyset$ and if $s \smallfrown \langle u \rangle \in T$ then $(s \smallfrown \langle u \rangle)' = s' \smallfrown \langle i \rangle$ such that $s \smallfrown \langle u \rangle$ is the i -th immediate successor of s in T . Note that H is Σ_1 and one-one on T . Now the set of all sequences t such that $lh(t) = x$ and

each member of t is $< a$] is bounded by some b ; by $S\Sigma_1$

$$(\exists d)(\forall t < b)[(\exists s \in T)(H(s) = t) \rightarrow (\exists s < d)(s \in T \& H(s) = t)].$$

Evidently, d is the desired bound.

(2) Given x , let c be a bound for all $s \in T$ of length x . Assume that all T_s ($s \in T, lh(s) = x$) are bounded. Then

$$(\forall s < c)(\exists q)[s \in T \& lh(s) = x \rightarrow (\forall t \in T)(t \supseteq s \rightarrow t < q)].$$

By $B\Sigma_2$ we obtain

$$(\exists q)(\forall s < c)[s \in T \& lh(s) = x \rightarrow (\forall t \in T)(t \supseteq s \rightarrow t < q)];$$

thus

$$(\exists q)(\forall t \in T)(t < q),$$

a contradiction. □

1.15 Definitions ($I\Sigma_1$). Now let $F : [X]^u \rightarrow a$, $u \geq 3$, $X \Delta_1$ and unbounded, $F\Delta_1$. An increasing sequence s of elements of X is *prehomogeneous* if for each $i_1 < \dots < i_u < i_{u+1} < lh(s)$ we have $F((s)_{i_1}, \dots, (s)_{i_{u-1}}, (s)_{i_u}) = F((s)_{i_1}, \dots, (s)_{i_{u-1}}, (s)_{i_{u+1}})$, i.e. the value does not depend on the last argument. If $lh(s) \geq u - 1$ and $s \frown \langle q \rangle$ is prehomogeneous, then the *colour* of $\max(s)$ in $s \frown \langle q \rangle$ is the finite mapping associating to each increasing sequence $i_1 < i_2 < \dots < i_{u-2} < lh(s) - 1$ the value $F((s)_{i_1}, \dots, (s)_{i_{u-2}}, \max(s), q)$.

The definition of a minimal prehomogeneous extension is as above; s is h.m.p.h. if *either* $lh(s) < u - 1$ and s consists of the first $lh(s)$ elements of X or $lh(s) \geq u$, s begins by the first $(u - 1)$ elements of X and for each i between $u - 1$ and $lh(s) - 1$, $s \upharpoonright (i + 1)$ is a minimal prehomogeneous extension of $s \upharpoonright i$.

1.16 Lemma ($I\Sigma_1$). Let $F : [X]^u \rightarrow a$ be as in 1.15 and let T be the set of all h.m.p.h. sequences. Then T is Δ_1 and is an unbounded finitely branching tree.

Proof. Generalize the proof of 1.13 (but drop narrow branching; finite branching is evident). □

1.17 Lemma ($B\Sigma_{m+1}$). Let $u \geq 2$ and let $F : [X]^u \rightarrow a$ be low Δ_{m+1} , X low Δ_{m+1} and unbounded. Then the set of all h.m.p.h. sequences is a low Δ_{m+1} finitely branching unbounded tree.

Hint. Relativize the above.

1.18 Lemma. $B\Sigma_2 \vdash \omega \rightarrow (\omega)_{<\omega}^1(\text{low } \Delta_2, \text{low } \Delta_2)$, i.e. if X is low Δ_2 unbounded and $F : [X] \rightarrow a$ then there is an $i < a$ such that $F^{-1}(i)$ is unbounded.

Proof. Assume the contrary, i.e. $(\forall i < a)(\exists y)[(\forall u)(F(u) = i \rightarrow u < y)]$. Since F is low Δ_2 , the formula in [...] is Δ_2 and by $B\Sigma_2$ we get $(\exists y)(\forall i < a)[(\forall u)F(u) = i \rightarrow u < y]$, i.e. $(\exists y)(\forall u)(u \in X \rightarrow u < y)$, which is contradiction. $\square$

1.19 Proof of Theorem 1.5. By induction on n . For $n = 1$ see 1.18 (and the obvious relativization). Let $F : [X]^n \rightarrow a$ be as assumed. By 1.17 take the tree T of all h.m.p.h. sequences; it is low Δ_{m+1} and, by 1.3.10 (5) has a low Δ_{m+2} unbounded branch. The branch defines a low Δ_{m+2} unbounded prehomogeneous set Y and F defines on $[Y]^{n-1}$ a function $G : [Y]^{n-1} \rightarrow a$; G is low Δ_{m+2} . By the induction hypothesis, G has an unbounded homogeneous set Z which is low $\Delta_{m+1+n-1}$, i.e. low Δ_{m+n} . Z is homogeneous also for F . $\square$

1.20 Proof of Theorem 1.7. (1) The initial case for $n = 1$ is

$$I\Sigma_m \vdash (\forall q)(\forall z)(\omega \rightarrow (q)_z^1(LL_m));$$

due to relativization, LL_m may be replaced by Δ_m . We also take $m = 1$, i.e. we prove the following: if $X \in \Delta_1$ is unbounded and $F \in \Delta_1$ maps X into $(< z)$ then there is a relatively large set a of cardinality at least q such that F is constant on a . But this is easy: By $S\Sigma_1$ find a $b \in X$ such that for each colour $i < z$, $(\exists x \in X)(F(x) = i)$ implies $(\exists x < b)(x \in X \& F(x) = i)$. Then let X_0 be the set of first $(b+1) * z$ elements of X and let j be a colour such that $a = \{x \in X_0 \mid x \text{ has the colour } j\}$ has more than b elements. Since $\min(a) \leq b$, a is the desired relatively large homogeneous set.

The induction step is now analogous to the induction step in 1.19 but instead of low Δ_{m+1} and low Δ_{m+2} one works with LL_m and LL_{m+1} .

(2) First let us prove $I\Sigma_1 \vdash \omega \rightarrow (q)_k^2(LL_1)$. It is enough to replace LL_1 by Δ_1 and then relativize.

Assume k standard to be given; we proceed in $I\Sigma_1$. Let $X \in \Delta_1$ be unbounded, $F : [X]^2 \rightarrow k$, $F \in \Delta_1$, q arbitrary. Let T be the tree of all h.m.p.h. sequences; it is unbounded, k -branching and Δ_1 . (See 1.13.) By 1.14 (1), for each x there is an upper bound b for all elements $s \in T$ of length $lh(s) \leq x$. By LII_1 , there is a least such b ; call it $H(x)$ and observe that H is $\Sigma_0(\Sigma_1)$.

Take our q and put $G(x) = H(x) * (k+1)$ and $r = G^{k+1}(q)$. (Here we use the fact that k is standard; we may iterate G $(k+1)$ -many times.) Clearly, T has elements of arbitrary length; fix an $s \in T$ such that $lh(s) = r$. For $i = 0, \dots, k+1$, let s_i be the initial segment of s having the length $G^i(q)$. Thus if $r_i = lh(s_i)$ we get $r_{i+1} = G(r_i) = H(r_i) * (k+1)$. Assign colours to elements of s in the usual way and let $col(s_i)$ be the set of colours of elements of s_i . Pick an i such that $col(s_i) = col(s_{i+1})$ and let Z be the set of elements of s_{i+1} . The cardinality of X is $H(r_i) * (k+1)$, therefore for

some colour j , the set $a = \{x \in Z \mid x \text{ has the colour } j\}$ has a cardinality bigger than $H(r_i)$. Now $H(r_i)$ is the maximum of elements on the level r_i , thus $H(r_i) \geq s_i \geq \max(s_i)$ (maximum of elements of s_i). And since $\text{col}(s_i) = \text{col}(s_{i+1})$ we get $\min(a) \leq \max(s_i) \leq H(r_i) < \text{card}(a)$. This proves our claim. The induction step is as above. $\square$

1.21 Proof of Theorem 1.9. (a) Assume that x, z, q are such that for no y we have $[x, y] \rightarrow (q)_z^n$. Thus for each y there is a *counter-example*-mapping $f : [x, y]^n \rightarrow z$ with no homogeneous relatively large set having at least q elements. Assume we have fixed a Δ_1 enumeration of $[x, \infty]^n = \bigcup_{y > x} [x, y]^n$ by all numbers such that for each y , $[x, y]^n$ forms an initial segment of length d_y . Then each counterexample is coded by a sequence s of length d_y such for each $i < d_y$, $(s)_i < z$. The set of all counterexamples determines naturally a Δ_1 tree T which is Δ_1 -estimated and unbounded; by the low basis theorem (in IS_1) it has a LL_1 unbounded branch. This branch naturally determines a LL_1 mapping $F : [V]^n \rightarrow z$ (where V is the set of all numbers) with no relatively large homogeneous set having $\geq q$ elements. But this contradicts 1.7 (1).

(b) Replace n by $(n + 1)$, take k standard and apply 1.7. $\square$

(c) Proofs (of 1.6, 1.8, 1.10)

It remains to prove theorems 1.6, 1.8 and 1.10. The proofs depend neither on the above proofs nor on each other.

1.22 Remark. Observe that for $m, k > 1$, the following are equivalent over IS_1 :

$$\begin{aligned}\omega &\rightarrow (\omega)_{<\omega}^1(\Delta_m, \Delta_m) \\ \omega &\rightarrow (\omega)_{<\omega}^1(\Delta_m, \Delta_{m+k})\end{aligned}$$

(since for a $F : X \rightarrow a$ maximal homogeneous sets are just sets $F^{-1}(i)$, $i < a$).

1.23 Lemma. For $m \geq 1$,

$$(*) \quad IS_1 + \omega \rightarrow (\omega)_{<\omega}^1(\Delta_m, \Delta_m) \vdash B\Sigma_{m+1}.$$

Proof. By 1.2.23, $B\Sigma_{m+1}$ may be replaced by $R\Pi_{m-1}$. The proof is by induction on m . Let $m = 1$. Let θ be Π_0 and assume $(Cx)(\exists y < a)\theta(x, y)$. Let $\theta'(x, y) \equiv \theta(x, y) \& (\forall y' < y) \neg \theta(x, y')$ (minimal selector); then θ' is Π_0 , and defines a function $F(x) = i \equiv \theta'(x, y) \& y < a$. F is Π_0 , $\text{dom}(F)$ is Π_0 and unbounded; by $\omega \rightarrow (\omega)_a^1(\Delta_1, \Delta_1)$ we get an $i < a$ such that $F^{-1}(i)$ is unbounded, thus $(Cx)\theta(x, i)$. This proves $R\Pi_0$.

Now assume (*) for m and prove it for $m + 1$. Thus assume $\omega \rightarrow (\omega)_{<\omega}^1(\Delta_{m+1}, \Delta_{m+1})$. Then $\omega \rightarrow (\omega)_{<\omega}^1(\Delta_m, \Delta_m)$, therefore by the induction hypothesis we have $B\Sigma_{m+1}$. We want to prove $R\Pi_m$. Let θ be Π_m and assume $(Cx)(\exists y < a)\theta$. Define θ' as above; by $B\Sigma_{m+1}$ θ' is Δ_{m+1} and the rest is as above. This proves the lemma. $\square$

1.24. Till now we have investigated the combinatorial relation $\omega \rightarrow (\omega)_{<\omega}^n(\Delta_i, \Delta_j)$ (defined in 1.4, cf. 1.5). Denote this relation briefly by $Arrow(n, i, j)$. Let us now consider an apparently weaker partition relation, denoted by

$$\omega \rightarrow^0 (\omega)_{<\omega}^n(\Delta_i, \Delta_j) \text{ or briefly } Arrow^0(n, i, j)$$

(thus $\rightarrow$ replaced by $\rightarrow^0$) whose definition results from the definition of $Arrow(n, i, j)$ by assuming X to be just the whole universe V , thus:

For each Δ_1 function $F: [V]^n \rightarrow z$ (where z is any number) there is a Δ_j unbounded homogeneous set.

First consider the case $n = 1$. Evidently, in $I\Sigma_m$ ($m \geq 1$) we have $Arrow(1, m, m) \equiv Arrow^0(1, m, m)$, since each Δ_m unbounded set is isomorphic with V by a Δ_m mapping (cf. I.2.65). We prove even more:

1.25 Lemma. $I\Sigma_1$ proves the equivalence of $Arrow(1, m, m)$ and $Arrow^0(1, m, m)$.

Proof. By induction on m , show $I\Sigma_1 + Arrow^0(1, m, m) \vdash I\Sigma_m$. Assume this for m and work in $(I\Sigma_1 + Arrow^0(1, m + 1, m + 1))$. By the induction hypothesis we have $I\Sigma_m$, thus $Arrow(1, m, m)$ and by 1.22, $B\Sigma_{m+1}$. Given a non-empty Σ_{m+1} set X such that $x \in X \equiv (\exists y)\theta(x, y)$ for some Π_m -formula θ , and an $a \in X$, define $F(s) = \min\{x \leq a \mid (\exists y < s)\theta(x, y)\}$ if this set is non empty, $= a + 1$ otherwise.

By $B\Sigma_{m+1}$, $i = F(s)$ is Δ_{m+1} (and total) and by $Arrow^0(1, m + 1, m + 1)$, there is an $i < a + 2$ such that $F^{-1}(i)$ is unbounded. This i is $\min X$. This proves $L\Sigma_{m+1}$ and hence $I\Sigma_{m+1}$. $\square$

1.26 Lemma. For $n \geq 2$, $m \geq 1$, $I\Sigma_m$ proves the following:

$$Arrow^0(n, m, n + m) \text{ implies } Arrow^0(n - 1, m + 1, n + m).$$

Hint: Let $F: [V]^{n-1} \rightarrow a$ be Δ_{m+1} ; by the limit theorem I.3.2 let $F(x) = \lim_s G(x, s)$ for a Δ_m function G . We may assume $G: [V]^n \rightarrow z$. An unbounded homogeneous set for G is homogeneous for F as well.

1.27 Theorem. Over $I\Sigma_1$, the following are equivalent ($n, m \geq 1$)

(i) $B\Sigma_{n+m}$

- (ii) $Arrow(n, m, m+n)$, i.e.
 $\omega \rightarrow (\omega)_{<\omega}^n(\Delta_m, \Delta_{m+n})$
- (iii) $Arrow^0(n, m, m+n)$, i.e.
 $\omega \rightarrow^0 (\omega)_{<\omega}^n(\Delta_m, \Delta_{m+n})$.

Proof. The only implication to be proved is (iii) $\rightarrow$ (i); but for $n = 1$ it follows by 1.25 and 1.23 and for $n > 1$ it follows using 1.26: Indeed assume (iii) for some $n \geq 1$ (and all m – induction hypothesis) and let $Arrow^0(n+1, m, m+n+1)$. Then, in particular, $Arrow^0(1, m, m+n+1)$, thus $Arrow^0(1, m, m)$ and $B\Sigma_{m+1}$; hence we may apply 1.26 and get $Arrow^0(n, m+1, n+m+1)$ hence $B\Sigma_{n+m+1}$ by the induction hypothesis. $\square$

Clearly, Theorem 1.6 follows. Proofs of 1.8 and 1.10 will be sketchy; the reader may elaborate details as an exercise.

1.28 Theorem. For $m \geq 1$, $B\Sigma_{m+1}$ proves that there is a Δ_m mapping $F : [V]^2 \rightarrow 2$ (where V is the universe of all numbers) having no Σ_{m+1} o.t.u. set.

Hint: The proof in [Jockusch, 1972-JSL] (Theorem 3.1) formalizes easily and gives a Δ_m mapping $F : V^2 \rightarrow 2$ with no o.t.u. Δ_{m+1} homogeneous set. By I.3.24 F has no o.t.u. Σ_{m+1} homogeneous set.

1.29 Corollary. (1) For $m \geq 1$, $I\Sigma_{m+1}$ proves

$$\neg[\omega \rightarrow^0 (\omega)_2^2(\Delta_m, \Sigma_{m+1})].$$

(2) Theorem 1.8 follows.

Hint: $I\Sigma_{m+1}$ proves that a Σ_{m+1} set is o.t.u. iff it is unbounded, see I.3.23.

1.30. We sketch a proof of the finite Ramsey theorem. It uses the following lemma:

(*) For each $u, a, q \geq 1$ there is a y such that for each x of cardinality y and each $f : [x]^u \rightarrow a$ there is a *prehomogenous* sequence of length q .

Suppose that for given u, a, q (*) does not hold and consider the tree of counterexamples like in 1.21. It is unbounded and Δ_1 estimated; an infinite LL_1 branch determines an infinite LL_1 mapping $F : [V]^u \rightarrow a$ with no prehomogeneous sequence of length q . But this contradicts 1.15, 1.16.

Now let us prove the finite Ramsey theorem.

Let $ph(u, a, q)$ be the minimal y satisfying (*); ph is Δ_1 and total. Define

$$\begin{aligned} rms(1, a, q) &= aq \\ rms(u+1, a, q) &= ph(u+1, a, rms(u, a, q)). \end{aligned}$$

The function rms is Δ_1 , total, and it is easy to show by induction on u that for each x of cardinality $rms(u, a, q)$ and each $f : [x]^u \rightarrow a$ there is a homogeneous sequence s (of elements of x) such that $lh(s) = q$. $\square$

2. Instances of the Paris-Harrington Principle and Consistency Statements

(a) Introduction and Statement of Results

2.1 Introduction. In Sect. 1 we introduced the notion of a relatively large finite set (X is relatively large if $\min X < |X|$) and the "arrow" notation $[x, y] \rightarrow_*^u (q)_z^u$ (for each $f : [x, y]^u \rightarrow z$ there is a relatively large homogeneous set having at least q elements). We put

$$PH(u, z) \equiv (\forall x, q)(\exists y)([x, y] \rightarrow_*^u (q)_z^u);$$

the *Paris-Harrington principle* was defined as the statement $(\forall u, z)PH(u + 1, z)$. Write (PH) for the last statement and $(PH)_u$ for the formula $(\forall z)PH(u + 1, z)$. Paris and Harrington showed that PA proves

$$(PH) \equiv Con^*(PA^* + Tr(\Pi_1^*))$$

where in $Con^*(...)$, PA^* stands for the natural Δ_1 definition of PA and $Tr(\Pi_1^*)$ means the set of all true Π_1^* -sentences. It follows by Gödel's second incompleteness theorem that (PH) is unprovable in PA . As it was shown above (1.9) for each $n \geq 1$,

$$I\Sigma_n \vdash (PH)_{n-1} \quad (\text{i.e. } I\Sigma_n \vdash (\forall z)PH(\bar{n}, z))$$

and $I\Sigma_n$ proves all numerical instances of $(PH)_n$, i.e. for each k , $I\Sigma_n \vdash PH(\bar{n} + 1, \bar{k})$.

The question whether the formulas $(PH)_n$ are related to statements assuring the consistency of fragments of $PA(+Tr(\Pi_1))$ is answered as follows by Paris's beautiful refinement of the result of Paris and Harrington:

2.2 Theorem. $I\Sigma_1$ proves that, for each $u \geq 1$,

$$(PH)_u \equiv Con^*(I\Sigma_u^* + Tr(\Pi_1^*)).$$

The proof of this result is the main content of the present section.

2.3 Corollary. (1) For each $n \geq 1$, $(PH)_n$ is provable in $I\Sigma_{n+1}$ but not in $I\Sigma_n$.

(2) $I\Sigma_1$ proves that $(PH) \equiv \text{Con}(PA + \text{Tr}(\Pi_1))$.

(1) follows by Gödel's second incompleteness theorem, (2) is immediate from 2.2 (and the compactness theorem).

2.4 Discussion. Both $(PH)_n$ and $\text{Con}(I\Sigma_n + \text{Tr}(\Pi_n))$ are Π_2 -statements; thus we have a hierarchy of sentences (1) forming an increasing hierarchy (the n -th of them is provable in $I\Sigma_{n+1}$ but not in $I\Sigma_n$), (2) being syntactically simple (Π_2) and (3) having a well understood double meaning: (a) combinatorial (mathematical), an instance of the Paris-Harrington principle, and (b) logical (metamathematical) – claiming the consistency of $I\Sigma_n^* + \text{Tr}^*(\Pi_1)$, which is a certain reflection principle for $I\Sigma_n^*$ (as we shall see later).

Non-provabilities are negative results; but they follow immediately from the positive result 2.1 via Gödel's second incompleteness theorem so it is natural to mention them here.

Let us now present our general plan of the proof. In subsection (b) we prove some combinatorial facts related to $(PH)_u$ and as a by-product we find a simplified formulation of $(PH)_u$. In (c) we prove the implication $\text{Con}^*(I\Sigma_u^* + \text{Tr}^*(\Pi_1)) \rightarrow (PH)_u$. We shall follow the corresponding proof of $\text{Con}(PA + \text{Tr}(\Pi_1)) \rightarrow (PH)$ due to Paris and Harrington. Paris's proof of the former implication uses properties of α -large sets (α an ordinal) elaborated in the next section. The subsections (d)–(e) contain a proof of $(PH)_u \rightarrow \text{Con}(I\Sigma_u^* + \text{Tr}(\Pi_1))$, together with various auxiliary things possibly useful elsewhere.

(b) Some Combinatorics

Recall that $PH(u, z)$ means

$$(\forall x, q)(\exists y)([x, y] \xrightarrow{*} (q)_z^u).$$

Note the obvious monotonicities:

2.5 Lemma ($I\Sigma_1$). If $[x, y] \xrightarrow{*} (q)_z^u$ and $x' \leq x$, $q' \leq q$, $z' \leq z$ and $y' \geq y$ then $[x', y'] \xrightarrow{*} (q')_{z'}^u$.

We are going to prove two results:

2.6 Theorem. $I\Sigma_1$ proves that, for each $u \geq 1$,

$$(PH)_u \equiv (\forall z)PH(u+1, z) \equiv (\forall z)(\exists y)([0, y] \xrightarrow{*} (u+2)_z^{u+1}).$$

2.7 Theorem. $I\Sigma_1$ proves that for each $u \geq 1$, $(\forall z)(\exists y)([0, y] \xrightarrow{*} (u+2)_z^{u+1})$ implies that for each z there is a y such that for each $f : [0, y]^{u+1} \rightarrow z$ there

is an H homogeneous for f and satisfying the following:

$$z \leq \min H \leq 2^{\min(H)} \leq |H|.$$

2.8 Remark. We are going to prove 2.7; our proof follows an analogous proof from [Paris-Harrington]. Then we indicate how to prove 2.6 by the same methods.

The following lemmas are proved in IS_1 :

2.9 Lemma. Let $f : [0, b]^e \rightarrow c$. $H \subseteq [0, b]$ is homogeneous for f iff each $(e+1)$ -element subset of H is.

2.10 Lemma. Let $f_i : [0, b]^e \rightarrow c_i$, $i = 1, \dots, k$ and let $f(\mathbf{x}) = \langle f_1(\mathbf{x}), \dots, f_k(\mathbf{x}) \rangle$. Then $f : [0, b]^e \rightarrow \prod c_i$ and $H \subseteq [0, b]$ is homogeneous for f iff it is homogeneous for each f_i .

2.11 Lemma. Let $f : [0, b]^e \rightarrow c$. Then there is an $f' : [0, b]^{e+1} \rightarrow c+1$ such that a set $H \subseteq [0, b]$, $|H| > e+1$, is homogeneous for f iff it is homogeneous for f' .

Proof. For $\mathbf{x} \in [0, b]^{e+1}$ put $f'(\mathbf{x}) = 0$ iff $\mathbf{x}$ is homogeneous for f , $f'(\mathbf{x}) = f(x_0, \dots, x_{e-1}) + 1$ otherwise. If H is homogeneous for f then clearly $f'(\mathbf{x}) = 0$ for each $\mathbf{x} \in [H]^{e+1}$. Conversely, let H be homogeneous for f' ; we prove that the value of f' on $[H]^{e+1}$ is 0, which implies that H is homogeneous for f . Let $\mathbf{x}$ be the least $(e+1)$ -tuple in H and $f'(\mathbf{x}) = i = 1 + f(x_0, \dots, x_{e-1})$. Let $y > x_{e+1}$ be another element of H ; for each $\mathbf{u} \in [\mathbf{x}]^e$,

$$f'(\mathbf{u}, y) = f'(x_0, \dots, x_e) = 1 + f(x_0, \dots, x_{e-1}) = 1 + f(\mathbf{u}),$$

thus $\mathbf{x}$ is homogeneous for f , contrary to our assumption. □

2.12 Remark. One can construct an $f' : [0, b]^{e+1} \rightarrow 1 + 2\sqrt{c}$ by refining the construction.

2.13 Lemma. For each b , there is an $f : [0, b]^2 \rightarrow 8$ such that, for each H relatively large and homogeneous for f ,

$$x, y \in H \text{ and } x < y \text{ implies } 2^x < y.$$

Proof. Let $f_0(x, y) = 0$ if $2x < y$, $= 1$ o.w.
 $f_1(x, y) = 0$ if $x^2 < y$, $= 1$ o.w.
 $f_2(x, y) = 0$ if $2^x < y$, $= 1$ o.w.

Let f combine all these in accordance with Lemma 2.10 and let H be homogeneous and relatively large. Let $a = \min H < |H|$, $e = \max H$, thus $2a \leq e$, $f_0(a, e) = 0$ and therefore $f_0(x, y) = 0$ for each $(x, y) \in [H]^2$. Thus $a^2 < e$, $f_1(a, b) = 0$ and therefore $f_1(x, y) = 0$ for each $(x, y) \in [H]^2$. Similarly, $2^a < e$ and $2^x < y$ for each $(x, y) \in [H]^2$. $\square$

2.14 Lemma. For each b, e there is an $f : [0, b]^e \rightarrow e + 6$ such that, for each H relatively large, homogeneous for f and each $(x, y) \in [H]^2$ such that $x < y$, we have $2^x < y$.

By Lemmas 2.13 and 2.11.

2.15 Lemma. For each b, e, c there is an $f : [0, b]^e \rightarrow c + 1$ such that for each H homogeneous for f and such that $|H| \geq e + 1$ we have $\min H \geq c$.

Proof. Let $f(x_1, \dots, x_e) = \min(x_1, c)$. $\square$

2.16 Lemma. For each $f : [a, b]^e \rightarrow c$ there is an $\hat{f} : [0, b]^e \rightarrow c(c + 1)(e + 6)$ such that if there is an H homogeneous for $\hat{f}$ and relatively large then there is an H' homogeneous for f such that

$$c \leq \min H' \leq 2^{\min H'} < |H'|.$$

Proof. Using Lemma 2.15 and Lemma 2.10, replace f by $f_0[0, b]^e \rightarrow c(c + 1)$ such that each H homogeneous for f_0 is homogeneous for f and satisfies $\min H \geq c$. Let $\log x$ be the maximal u such that $2^u \leq x$, let $\log(x_1, \dots, x_n)$ be $(\log(x_1), \dots, \log(x_n))$. Define

$$\begin{aligned} f_1(\mathbf{x}) &= f_0(\log(\mathbf{x})) \text{ for } \mathbf{x} \in [0, b]^e. \\ p : [0, b]^e &\rightarrow e + 6 \text{ from Lemma 2.14} \end{aligned}$$

Let $\hat{f}(\mathbf{x})$ combine f_1, p and let H be relatively large, homogeneous for $\hat{f}$. Then H is homogeneous for f_1 , $\min H \geq c$ and we have $2^x < y$ for $(x, y) \in [H]^2$.

Let $H' = \{\log x \mid x \in H\}$; we have $|H'| = |H|$, H' is homogeneous for f_0 , $\min H' = \log(\min H)$, thus $2^{\min H'} < |H'|$ as desired. $\square$

2.17 Remark. Theorem 2.7 follows directly. To prove 2.6 assume $(\forall z)(\exists y)([0, y] \rightarrow_* (u + 2)_z^{u+1})$ and let x, q, z be given; let $z' = z \cdot \max(x, q)$ and let y be such that $[0, y] \rightarrow_* (u + 2)_{z'}^{u+1}$. Assume $f : [x, y]^{u+1} \rightarrow z$; extend f arbitrarily to an $f_0 : [0, y]^{u+1} \rightarrow z$ and combine it with $f_1 : [0, y]^{u+1} \rightarrow \max(x, q)$ such that each H homogeneous for f_1 satisfies $\min(H) \geq x, q$. Let $\hat{f}$ be the resulting function and let H be homogeneous for $\hat{f}$, H relatively large, $|H| > u + 1$. Then $\min H \geq x, q$ and H is homogeneous for f .

(c) Proof of $Con^*(I\Sigma_u^* + Tr(\Pi_1^*)) \rightarrow (PH)_u$ (for $u \geq 1$)

2.18 Proof. Recall 1.9: there we proved that for each $n \geq 1$ and for each $x \geq 1$, $I\Sigma_n$ proves $PH(n+1, k)$ (i.e. proves $(\forall x, q)(\exists y)([x, y] \rightarrow_* (q)_k^{n+1})$). The proof of this fact formalizes in $I\Sigma_1$, as an easy inspection shows, so that we have the following:

$$I\Sigma_1 \vdash (\forall z \geq 1)(\forall u \geq 1) Pr_{I\Sigma_u}^*(PH(u+1, z)) \quad (')$$

Now work in $I\Sigma_1 + Con^*(I\Sigma_u^* + Tr(\Pi_1^*))$. The added axiom can be evidently reformulated as saying that each Σ_1 -sentence provable in $I\Sigma_u$ is true (otherwise its negation would be a true Π_1 -sentence inconsistent with $I\Sigma_u$). Now take any x, z, q and observe that, by $(')$, $I\Sigma_u^*$ proves $(\exists y)([x, y] \rightarrow_* (\dot{q})_z^{u+1})$, which is a Σ_1^* -sentence. Thus this sentence is true (in the sense of satisfaction of Σ_1^* -sentences). But then, by the "it's snowing"-it's snowing lemma, we get $(\exists y)([x, y] \rightarrow_* (q)_z^u)$; we have proved $(\forall z)PH(u+1, z)$.

(d) Strong Indiscernibles

Recall the Σ'_n -formulas introduced in Chap. I, Sect. 2 (e).

2.19 Definition. For each Σ'_n -formula φ having the form

$$(\exists y_1)(\forall y_2) \dots \varphi_0(\mathbf{x}, \mathbf{y})$$

let $\varphi \upharpoonright$ be the Σ'_n -formula

$$(\exists y_1 \leq v_1)(\forall y_2 \leq v_2) \dots \varphi_0(\mathbf{x}, \mathbf{y})$$

where $v_1, \dots, v_n$ are variables not occurring in φ ; they are called the *designated variables* of $\varphi \upharpoonright$. Let $(\Sigma'_n) \upharpoonright$ denote the set of all $\varphi \upharpoonright$ for $\varphi \in \Sigma'_n$.

2.20 Observation. Definition 2.19 is meaningful in $I\Sigma_1$; thus in $I\Sigma_1$ we have, for each u , the Δ_1 -set of all $(\Sigma'_u \upharpoonright^*)$ -formulas. Moreover, since $(\Sigma'_u \upharpoonright^*)$ -formulas are particular Σ_0^* -formulas and therefore we have a Δ_1 satisfaction for all $(\Sigma'_u \upharpoonright^*)$ -formulas (with arbitrary u); we denote it occasionally by $\models$.

2.21 Definition (in $I\Sigma_1$). A finite set $B = \{b_i \mid i < \lambda\}$ (increasing enumeration) is a set of *strong indiscernibles* for $(\Sigma'_u \upharpoonright^*)$ -formulas^{*} if for each $i < \lambda$ we have the following:

For each $(\Sigma'_u \upharpoonright^*)$ -formula^{*} $\varphi(\mathbf{x}, \mathbf{v})$ ($\mathbf{v}$ designated), such that $\varphi(\mathbf{x}, \mathbf{v}) < i$, each tuple $\mathbf{p}$ of possible meanings of $\mathbf{x}$, all $\leq b_i$, and each pair $\mathbf{b}, \mathbf{b}' \in (B \setminus [0, b_i])^u$ of increasing u -tuples of elements of B bigger than b_i ,

$$(*) \quad \models \varphi(\mathbf{p}, \mathbf{b}) \equiv \varphi(\mathbf{p}, \mathbf{b}').$$

(Remember that in the last equivalence, φ must be sufficiently small ($\leq i$), the parameters $\mathbf{p}$ must be sufficiently small ($\leq b_i$) and increasing u -tuples $\mathbf{b}, \mathbf{b}'$ of elements B sufficiently large (all elements $> b_i$).

2.22 Example. Let $u = 3$, let φ be

$$(\exists y_1 \leq v_1)(\forall y_2 \leq v_2)(\exists y_3 \leq v_3)\psi(x_1, x_2, y).$$

Assume $\varphi \leq i$; $p_1, p_2 \leq b_i$; $i < j < k < q$; $i < j' < k < q$. Then $(*)$ implies

$$\models (\exists y_1 \leq b_j)(\forall y_2 \leq b_k)(\exists y_3 \leq b_q)\psi(p_1, p_2, y)$$

iff

$$\models (\exists y_1 \leq b_{j'})(\forall y_2 \leq b_{k'})(\exists y_3 \leq b_{q'})\psi(p_1, p_2, y)$$

2.23 Theorem. IS_1 proves that, for each u , $(PH)_u$ implies the following: For each ν there is a set B of strong indiscernibles for $(\Sigma'_u \upharpoonright^*)$ -formulas such that $|B| = \nu$.

We prove this theorem in the present section. The next section is devoted to a proof of the fact that, for each u , the conclusion of 2.23 (existence of arbitrarily large sets of strong indiscernibles for $(\Sigma'_u \upharpoonright^*)$ -formulas^{*} implies $Con^*(IS'_u + Tr(\Pi_1^*))$.

For simplicity, we shall assume $u = 3$. But the method is perfectly general.

2.24 Conventions (only for this section). Define in IS_1 as follows: let $(\exists x \leq v)\varphi(x, \mathbf{z})$ be a Σ'_0 -formula^{*} such that v does not occur in φ ; let $\mathbf{p}$ be a tuple of possible meanings of $\mathbf{z}$. Then the element *defined* in $[0, d]$ by this formula with parameters $\mathbf{p}$ is the minimal $a \leq d$ such that $\models \varphi(a, \mathbf{p})$ (if there is such an a). Dually, the element *defined* in $[0, d]$ by $(\forall x \leq v)\varphi(x, \mathbf{z})$ is the minimal $a \leq d$ such that $\models \neg\varphi(a, \mathbf{p})$.

If d is a number and $b \subseteq [0, d]$ then $def_q(d, b)$ denotes the set of all elements of $[0, d]$ defined by formulas^{*} ψ of the above two forms such that $\psi < q$, with parameters from b . (In particular, you may use for ψ any $(\Sigma'_u \upharpoonright^*)$ -formula^{*} or $(\Pi'_u \upharpoonright^*)$ -formula^{*} ($u \geq 1$) w.r.t. its first designated variable.)

Let $\beta < \gamma < \delta < d$ be given. An increasing sequence $(a_q \mid q < \nu)$ of elements less than β is a *Paris sequence* (for β, γ, δ, d) if, for each $q < \nu - 1$,

- (1) $[a_{q+1}, \beta] \cap def_q(d, [0, a_q] \cup \{\gamma, \delta\}) = \emptyset$,
- (2) $[a_{q+1}, \gamma] \cap def_q(d, [0, a_q] \cup \{\delta\}) = \emptyset$,
- (3) $[a_{q+1}, \delta] \cap def_q(d, [0, a_q]) = \emptyset$.

2.25 Lemma (IS_1). A Paris sequence is a set of strong indiscernibles for $(\Sigma'_3 \upharpoonright^*)$ -formulas^{*}.

Proof. Let $\psi(\mathbf{w}) \equiv (\exists x)(\forall y)(\exists z)\varphi(x, y, z, \mathbf{w})$ where φ is Σ'_0 ; assume $(\psi \upharpoonright) < i < j < k < q < \nu$, $\mathbf{p} \leq a_i$. Then the following is true (in the sense of $\models$):

$$\begin{aligned} (\psi \upharpoonright)(\beta, \gamma, \delta, \mathbf{p}) &\equiv (\exists x \leq \beta)(\forall y \leq \gamma)(\exists z \leq \delta)\varphi(x, y, z, \mathbf{p}) \equiv \\ &\equiv (\exists x \leq a_j)(\forall y \leq \gamma)(\exists z \leq \delta)\varphi(x, y, z, \mathbf{p}) \end{aligned}$$

(since otherwise the smallest x such that $(\forall y \leq \gamma)(\exists z \leq \delta)\varphi(x, y, z, \mathbf{p})$ would be in $[a_j, \beta]$, i.e. in $\text{def}_i(d, [0, a_i] \cup \{\gamma, \delta\}) \cap [a_{i+1}, \beta]$, which contradicts (1)),

$$\equiv (\exists x \leq a_j)(\forall y \leq a_k)(\exists z \leq \delta)\varphi(x, y, z, \mathbf{p})$$

(since if we let x_0 be such that $x_0 \leq a_j$ and $(\forall y \leq a_k)(\exists z \leq \delta)\varphi(x_0, y, z, \mathbf{p})$ but not $(\forall y \leq \gamma)(\exists z \leq \delta)(\dots)$, then the minimal y such that $\neg(\exists z \leq \delta)\varphi(x_0, y, z, \mathbf{p})$ would lie in $[a_k, \gamma]$, hence in $\text{def}_j(d, [0, a_j] \cup \{\delta\}) \cap [a_{j+1}, \gamma]$, which contradicts (2))

$$\equiv (\exists x \leq a_j)(\forall y \leq a_k)(\exists z \leq a_q)\varphi(x, y, z, \mathbf{p})$$

(otherwise let $x_0 \leq a_j$ be minimal such that

$$(\forall y \leq a_k)(\exists z \leq \delta)\varphi(x_0, y, z, \mathbf{p})$$

and take a $y_0 \leq a_k$ such that

$$\neg(\exists z \leq a_q)\varphi(x_0, y_0, z, \mathbf{p});$$

then the minimal z such that $\varphi(x_0, y_0, z, \mathbf{p})$ would lie in $[a_q, \delta]$, thus in $\text{def}_k(d, [0, a_k]) \cap [a_{k+1}, \delta]$, which contradicts (3)).

Thus

$$\models (\psi \upharpoonright)(\beta, \gamma, \delta, \mathbf{p}) \equiv (\psi \upharpoonright)(a_i, a_j, a_k, \mathbf{p})$$

for all $i < j < k$ satisfying our condition, which shows that the Paris sequence $(a_q \mid q < \nu)$ is a sequence of strong indiscernibles for $(\Sigma'_3 \upharpoonright)$ -formulas. $\square$

To complete the proof of Theorem 2.23, it remains to prove the following

2.26 Lemma ($I\Sigma_1$). For each u , $(PH)_u$ implies the existence of a Paris sequence of an arbitrary length.

Proof (for $u = 3$). Let the desired length $\nu \geq 5$ of a Paris sequence be given. We assume $(PH)_3$, and use 2.7. Take a sufficiently large c (w.r.t. ν ; it turns out that $c = 2^\nu$ is sufficient) and let d be such that for each $f : [0, d]^4 \rightarrow c$ there is a homogeneous H such that $c \leq \min H \leq 2^{\min H} \leq |H|$.

For each $\alpha < \beta < \gamma < \delta < d$ define (cf. (1) above)

$$a_0 = 0$$

$$a_{q+1} = \max[\text{def}_q(d, [0, a_q] \cup \{\gamma, \delta\}) \cap [0, \beta]] + 1$$

($q = 0, 1, \dots, \nu - 1$).

Thus $[a_{q+1}, \beta] \cap \text{def}_q(d, [0, a_q] \cup \{\gamma, \delta\}) = \emptyset$. Observe that if $a_q < \beta$ then $a_q < a_{q+1}$. We want to find $\alpha < \beta < \gamma < \delta < d$ such that the corresponding sequence of a_q 's satisfies the following for each $q < \nu - 1$.

- (1') $a_{q+1} \leq \alpha$, i.e. $[\alpha, \beta] \cap \text{def}_q(d, [0, a_q] \cup \{\gamma, \delta\}) = \emptyset$,
 (2') (2) , i.e. $[\beta, \gamma] \cap \text{def}_q(d, [0, a_q] \cup \{\gamma\}) = \emptyset$,
 (3') (3) , i.e. $[\gamma, \delta] \cap \text{def}_q(d, [0, a_q]) = \emptyset$.

Define the function $F : [0, d]^4 \rightarrow c$ as follows: for each $\langle \alpha, \beta, \gamma, \delta \rangle \in [0, d]^4$ let $(a_q \mid q < \nu)$ be the sequence defined above and put

$$\begin{aligned} F(\alpha, \beta, \gamma, \delta) &= (\min q < \nu)([\alpha, \beta] \cap \text{def}_q(d, [0, a_q] \cup \{\gamma, \delta\}) \neq \emptyset) \\ &\quad \text{if there is such a } q, \text{ else} \\ &= c/4 + (\min q < \nu)([\beta, \gamma] \cap \text{def}_q(d, [0, a_q] \cup \{\gamma\}) \neq \emptyset) \\ &\quad \text{if there is such a } q, \text{ else} \\ &= c/2 + (\min q < \nu)([\gamma, \delta] \cap \text{def}_q(d, [0, a_q]) \neq \emptyset) \\ &\quad \text{if there is such a } q, \text{ else} \\ &= 3c/4 + 1 \end{aligned}$$

Evidently, $F : [0, d]^4 \rightarrow c$; if we prove that there is a homogeneous H such that the common value of F on $[H]^4$ is $3c/4 + 1$, then each quadruple $\langle \alpha, \beta, \gamma, \delta \rangle \in H$ determines a Paris sequence of length ν .

Now let H be homogeneous for F and such that $c \leq \min H \leq 2^{\min H} < |H|$; let $\{h_i \mid i < e\}$ be its increasing enumeration. First assume that the common value of F on H^4 is $q < c/4$. Then for $h_i < h_j < h_k < h_m$ from H we have

$$[h_i, h_j] \cap \text{def}_q(d, [0, a_q] \cup \{h_k, h_m\}) \neq \emptyset$$

and since $[h_i, h_j] \cap \text{def}_{q-1}(d, [0, a_{q-1}] \cup \{h_k, h_m\}) = \emptyset$ and $a_q = \max[\text{def}_{q-1}(d, [0, a_{q-1}] \cup \{h_k, h_m\}) \cap [0, h_j]] + 1$, we get $a_q \leq h_i$. Note that a_q depends on h_j, h_k, h_m but not on h_i ; we have $F(h_0, h_j, h_k, h_m) = F(h_0, h_j, h_k, h_m) = F(h_i, h_j, h_k, h_m) = q$ and we get $a_q \leq h_0$. Hence for all i we have

$$[h_i, h_{i+1}] \cap \text{def}_q(d, [0, h_0] \cup \{h_{e-2}, h_{e-1}\}) \neq \emptyset.$$

But we have $\leq q$ formulas each with $\leq q$ free variables and $h_0 + 3$ parameters, thus

$$|\text{def}_q(d, [0, h_0] \cup \{h_{e-2}, h_{e-1}\})| \leq (q+1) \cdot (h_0+3)^q \leq (h_0+3)^{(q+1)}$$

(since $q < \nu < c < h_0$). The last set must intersect each of $(|H| - 3)/2$ disjoint intervals $[h_{2i-1}, h_{2i})$ ($i = 1, \dots, e - 2$). From $2^{h_0} < |H|$ we get

$2^{h-2} < (2^{h_0} - 3)/2 < (|H| - 3)/2$; thus if we can prove $(h_0 + 3)^{(q+1)} \leq 2^{h_0-2}$, we have a contradiction. Now remember that we took $c = 2^\nu$. Thus $(h_0 + 3)^{(q+1)} \leq (h_0 + 3)^\nu$ and $(h_0 + 3) > h_0 \geq c = 2^\nu$. Hence it suffices to prove that $x \geq 2^\nu$ implies $x^\nu \leq 2^{x-5}$. But evidently this is true for $x = 2^\nu$ if $\nu \geq 5$ and therefore true for each $x \geq 2^\nu$ (if $\nu \geq 5$). Thus we get a contradiction and have excluded the first possibility in the definition of F .

Similarly we eliminate the second and third possibility. Take the second. We already know that $a_q < h_0$ for each q (since the first case does not occur). Assume $F(h_i, h_j, h_k, h_m) = c/4 + q$ for all respective h 's. Thus

$$\begin{aligned} [h_j, h_k] \cap \text{def}_q(d, [0, a_q] \cup \{h_m\}) &= \emptyset, \text{ thus} \\ [h_j, h_{j+1}] \cap \text{def}_q(d, [0_0] \cup \{h_{e-1}\}) &= \emptyset \end{aligned}$$

for all j (as above), which leads to a contradiction. The third case is analogous.

Thus the common value of F on $[H]^4$ is $c/4 + 1$ and therefore for each $\langle \alpha, \beta, \gamma, \delta \rangle \in [H]^4$ the corresponding sequence of a 's is a Paris sequence. This completes the proof of Lemma 2.26 and of the Theorem 2.23. $\square$

(e) Final Considerations

2.27 Recall Theorem I.4.37; it will be used to complete the proof of $(\forall m \geq 2) ((PH)_{m-1} \rightarrow \text{Con}^*(L\Pi'_{m-1} \bullet \cup \text{Tr}(\Pi'_1 \bullet)))$ in $I\Sigma_1$. (We use $m-1$ instead of u and $L\Pi'_{m-1}$ instead of $I\Sigma'_{m-1}$ to simplify our considerations.) Let S_0 be a finite set of closed instances of $Sk_0(L\Pi'_{m-1} \bullet \cup \text{Tr}(\Pi'_1 \bullet))$; assuming $(PH)_{m-1}$ we shall construct a Δ_1 -satisfaction $\models'$ for S_0 such that $\models'$ extends $\models$ (the usual Δ_1 satisfaction for Σ_0 -formulas) and $\models' S_0$. By I.4.37, this implies the desired consistency. The satisfaction $\models'$ is constructed using a sufficiently long set of strong $(\Sigma'_{m-1} \upharpoonright \bullet)$ -indiscernibles, guaranteed by 2.26. The definition follows.

2.28 Definition ($I\Sigma_1$). Let Φ be $(Q_1x_1) \dots (Q_kx_k)\varphi(\mathbf{x}, \mathbf{y})$, $\varphi \in \Sigma_0$. Recall the meaning of $(\Phi \upharpoonright \mathbf{z})(\mathbf{x}, \mathbf{y})$, namely

$$(Qx_1 \leq z_1) \dots (Q_kx_k \leq z_k)\varphi(\mathbf{x}, \mathbf{y}).$$

Let $A = \{a_q \mid q < \nu\}$ be a finite set in its increasing enumeration. $(\Phi \upharpoonright a_q \rightarrow)$ obviously means the result of substituting $a_q, a_{q+1} \dots$ for $z_1, z_2 \dots$ into $(\Phi \upharpoonright \mathbf{z})$, i.e.

$$(Q_1x_1 \leq a_q) \dots (Q_kx_k \leq a_{q+k-1})\varphi(\mathbf{x}, \mathbf{y}).$$

Given $c_i, \mathbf{d}$, assume that q is the least number such that

- (1) $(\Phi \upharpoonright) \leq q$,
- (2) $\leftarrow c_{i-1}, \mathbf{d} \leq a_q$,

$$(3) \ q + k \leq \nu.$$

If $Q_i = \exists$ put

$$f_i^\Phi(\leftarrow c_{i-1}, \mathbf{d}) = (\min c_i \leq a_{q+1})[\models (\Phi^{(i)} \upharpoonright a_{q+2} \rightarrow)(\leftarrow c_i, \mathbf{d})].$$

In all other cases put $f_i^\Phi(\leftarrow c_i^*, \mathbf{d}) = 0$.

Thus we have interpreted all the function symbols of $Sk_0(\Phi)$ by Δ_1 functions (dependent on A). Similarly for a finite set of formulas instead just one. This determines, in the usual way, a Δ_1 satisfaction $\models'$ for any finite set of formulas of the form $Sk_0(\Phi)$ (cf. I.4.14).

2.29 Theorem (IS_1). Let $m \geq 2$ and $(PH)_{m-1}$. For each finite set S_0 of closed instances of $Sk_0(L\Pi'_{m-1} \bullet \cup Tr(\Pi'_1 \bullet))$ there is a ν such that if A is a set of strong indiscernibles for $(\Sigma'_{m-1} \upharpoonright \bullet)$ -formulas of the cardinality ν and $\models'$ is the satisfaction for S_0 given by definition 2.28, then $\models' S_0$.

2.30 Corollary. Theorem 2.2 follows.

Elaboration.

2.31 Lemma (IS_1). If Φ is Σ'_{m-1} and A is a set of strong $(\Sigma'_{m-1} \upharpoonright \bullet)$ -indiscernibles, then for any q satisfying 2.28 (1)–(3) we have

$$f_i^\Phi(\leftarrow c_{i-1}, \mathbf{d}) = (\min c_i \leq a_{q+1})[\models (\Phi^{(i)} \upharpoonright a_{q+2} \rightarrow)(\leftarrow c_i, \mathbf{d})].$$

(Obvious from the definition of strong indiscernibles.)

2.32 Lemma (IS_1). Let Φ be Σ'_{m-1} or Π'_{m-1} and let $\varphi(\mathbf{s}, \mathbf{u})$ be a closed instance of $Sk_0(\Phi)$. Let A be a set of strong $(\Sigma'_{m-1} \upharpoonright \bullet)$ -indiscernibles, $\models'$ the corresponding satisfaction and let q satisfy 2.28 (1)–(3) for $\mathbf{c} = V(\mathbf{s})$, $\mathbf{d} = V(\mathbf{a})$. (V is the interpretation of the term $\mathbf{s}$).

Then

$$\models' (\Phi \upharpoonright a_{q+1})(\mathbf{u}) \rightarrow \varphi(\mathbf{s}, \mathbf{u}).$$

Proof. As in the proof of I.4.37 prove

$$\models' (\Phi \upharpoonright a_{q+1}(\mathbf{u}) \rightarrow (\Phi^{(i)} \upharpoonright a_{q+1})(\leftarrow s_i, \mathbf{u})).$$

It suffices to show

$$(\text{'}) \quad \models' (\Phi^{(i)} \upharpoonright a_{q+1})(\leftarrow s_i, \mathbf{u}) \rightarrow (\Phi^{(i+1)} \upharpoonright a_{q+1})(\leftarrow s_{i+1}, \mathbf{u}).$$

But $\Phi^{(i)}(\leftarrow x_i, \mathbf{y})$ is $(Q_{i+1}x_{i+1})\Phi^{(i+1)}(\leftarrow x_{i+1}, \mathbf{y})$.

Consider the following two cases:

Case 1. $Q_{i+1} = \forall$. Since $\models' s_{i+1} \leq a_q$ (by (2)), we have

$$\begin{aligned} \models' (\Phi^{(i)} \upharpoonright a_{q+1})(\leftarrow s_i, \mathbf{u}) &\rightarrow (\forall x_{i+1} \leq a_{q+1})(\Phi^{(i+1)} \upharpoonright a_{q+2})(\leftarrow s_i, x_{i+1}, \mathbf{u}) \\ &\rightarrow (\Phi^{(i+1)} \upharpoonright a_{q+2})(\leftarrow s_{i+1}, \mathbf{u}) \\ &\rightarrow (\Phi^{(i+1)} \upharpoonright a_{q+1})(\leftarrow s_{i+1}, \mathbf{u}) \end{aligned}$$

(by indiscernibility).

Case 2. $Q_{i+1} = \exists$. Thus $s_{i+1} = F_{i+1}^\Phi(\leftarrow s_i, \mathbf{u})$. Similarly as above, but using also the definition of f_{i+1}^Φ , we have $\models' (\Phi^{(i)} \upharpoonright a_{q+1})(\leftarrow s_i, \mathbf{u}) \rightarrow (\exists x_{i+1} \leq a_{q+1})(\Phi^{(i+1)} \upharpoonright a_{q+2})(\leftarrow s_i, x_{i+1}, \mathbf{u}) \rightarrow (\Phi^{(i+1)} \upharpoonright a_{q+2})(\leftarrow s_{i+1}, \mathbf{u}) \rightarrow (\Phi^{(i+1)} \upharpoonright a_{q+1})(\leftarrow s_{i+1}, \mathbf{u})$. This completes the proof. $\square$

2.33 Proof of 2.29. Let a finite set $S_0 \subseteq \text{inst}(Sk_0(\Pi'_{m-1} \cup Tr(\Pi'_1)))$ be given. Let ν_0 be such that for each instance $\psi(s_1, \dots, s_k) \in S_0$ of the Skolemization of an axiom $\Psi \in L\Pi'_{m-1} \cup Tr(\Pi'_1)$ we have

$$(\Psi \upharpoonright) < \nu_0 \text{ and } s_1, \dots, s_k < \nu_0$$

and let $A = \{a_q \mid q < \nu_0 + 3m\}$ be a set of strong $(\Sigma'_{m-1} \upharpoonright^*)$ -indiscernibles. (As we shall see, if $\Psi \in L\Pi'_{m-1}$ then its "prenex normal form with bounded kernel" has $\leq 3m$ unbounded quantifiers.) Our aim is to show $\models' S_0$ for the Δ_1 -satisfaction given by A .

(1) Let $\Psi \in Tr(\Pi'_1)$, $\Psi = (\forall x)\psi(x)$ where ψ is bounded'. Let $\psi(s) \in S_0$; then $\models \psi(V(s))$, i.e. $\models' \psi(s)$.

(2) Now let Ψ be $L\text{-}\Phi$, where $\Phi(x_1)$ is a Σ'_{m-1} -formula $(\exists x_2) \dots (Q_m x_m) \varphi(x_1, \dots, x_m)$. Thus Ψ is the following:

$$(\forall x_1)[\Phi(x_1) \vee (\exists y_1 \leq x_1)(\neg \Phi(y_1) \& (\forall z_1 < y_1)\Phi(z_1))].$$

Hence an instance of $Sk_0(\psi)$ has the form

$$\varphi(s) \vee (t_1 \leq s_1 \& \neg \varphi(t) \& (r_1 < t_1 \rightarrow \varphi(r))), \text{ in short, } \hat{\psi}(s, t, r).$$

(Precise conditions on the form of the terms s, t, r will be considered later.)

Let q be minimal such that $(L\text{-}\Phi \upharpoonright) \leq q$ and $V(s), V(t), V(r) \leq a_q$.

We have two cases

(a) $\models' \varphi(s)$; then $\hat{\psi}(s, t, r)$ and we are done.

(b) $\models' \neg \varphi(s)$; then, by Lemma 2.32, we have $\models' (\neg \Phi \upharpoonright a_{q+1})(s_1)$.

Let e be the least number such that $\models' (\neg \Phi \upharpoonright a_{q+1})(e)$; assume $\mathbf{b} = V(s)$ (i.e. $b_1 = V(s_1)$ etc.).

(3) *Claim.* $V(t_1) = e$.

Indeed, for some $h \leq q$,

$$\begin{aligned}
 f_{m+1}^L(\mathbf{b}) &= (\min c_1 \leq a_h) \models \varphi(\mathbf{b}) \vee .c_1 \leq b_1 \ \& \ (\neg\Phi \upharpoonright a_{h+1})(c_1) \\
 &\quad \& \ (\forall z_1 < c_1)(\Phi \upharpoonright a_{h+m+1})(z_1) \\
 &= (\min c_1 \leq a_h) \models (\neg\Phi \upharpoonright a_{q+1})(c_1) \ \& \ (\forall z_1 < c_1)(\Phi \upharpoonright a_{q+1})(z_1) \\
 &= (\min c_1 \leq a_h) \models (\neg\Phi \upharpoonright a_{q+1})(y_1) \\
 &= e;
 \end{aligned}$$

thus $V(t_1) = e$.

(4) Now by Lemma 2.32, $\models (\neg\Phi \upharpoonright a_{q+1})(t_1)$ implies $\models \neg\varphi(t'_1, t'_2, \dots, t'_m)$ where t'_i is t_i for $i = 1$ or i even and is given by $\neg\Phi$ for i odd, $i > 1$. Our t_i are given by $L_{\neg\Phi}$.

(5) *Claim.* For $i = 1, \dots, m$, $\models t_i = t'_i$; thus $\models \neg\varphi(\mathbf{t})$.

For i even and $i = 1$ we trivially have $\models t_i = t'_i$; for i odd, $i > 1$ we proceed by induction. Now $t'_i = F_i^{\neg\Phi}(\leftarrow t_{i-1})$ and $t_i = F_{m+1}^L(s, t_{i-1})$. Assume $\mathbf{b} = V(\mathbf{s})$ and $\leftarrow c_{i-1} = V(\leftarrow t_{i-1}) = V(\leftarrow t'_{i-1})$ and compute:

$$\begin{aligned}
 f_{m+i}^L(\mathbf{b}, \leftarrow c_{i-1}) &= (\min c_i \leq a_p)(\varphi(\mathbf{b}) \vee c_1 \leq b_1 \ \& \ (\neg\Phi^{(i)} \upharpoonright a_{p+1})(\leftarrow c_i) \\
 &\quad \& \ (\forall z_1 < c_1)\Phi \upharpoonright a_p(z_1) \\
 &= (\min c_i \leq a_p)(\neg\Phi^{(i)} \upharpoonright a_{p+1})(\leftarrow c_i) \\
 &= (\min c_i \leq a_q)(\neg\Phi^{(i)} \upharpoonright a_{q+1})(\leftarrow c_i) \\
 &= f_i^{\neg\Phi}(\leftarrow c_{i-1}).
 \end{aligned}$$

Thus $V(t_i) = V(t'_i)$.

(6) Now take r_1 ; if $\models r_1 \geq t_1$ nothing need be proved. Thus assume $V(r_1) < V(t_1) = d_1$; then $\models \Phi \upharpoonright a_{q+1}(t_1)$, therefore $\models \varphi(r'_1, \dots, r'_m)$ where $r'_i = r_i$ for i odd and similarly to above we prove $\models r_i = r'_i$ for i even. We have proved

(7) *Claim.* Under the present assumption we have $\varphi(\mathbf{r})$. Thus we have proved $\hat{\psi}(\mathbf{s}, \mathbf{t}, \mathbf{r})$, which completes the proof. $\square$

3. Schwichtenberg-Wainer Hierarchy and α -large Sets

In the preceding two sections we studied instances of Paris-Harrington principle and showed (1) that the k -th instance $(PH)_k$ is provable in IS_{k+1} and (2) that, provably in IS_1 , $(PH)_k$ is equivalent to $Con^*(IS_k^* + Tr^*(\Pi_1))$. (Evidently, (2) implies (1) since IS_{k+1} proves the above consistency, cf. I.4.33–34,

but the explicit proof of $(PH)_k$ in $I\Sigma_{k+1}$ that we presented is of independent interest.)

Now we are going to study a different but related combinatorial principle and its instances. We shall call it (W) or *the principle of α -large intervals*. For background see bibliographical remarks; we shall present the principle and relate its instances to instances of Paris-Harrington principle. This will be done in the following steps: (a) we introduce ordinals in $I\Sigma_1$ and derive their important properties, (b) we show which induction is sufficient to get enough induction for ordinals, (c) we introduce and study α -large sets, and (d) we define the principle (W) and relate instances of (W) to instances of Paris-Harrington principle. Note that results of this section may be used to get the characterization of functions provably total in $I\Sigma_k$ (and PA) using model theoretic means; this will be done in Chap. IV.

(a) Ordinals in $I\Sigma_1$

3.1. We are going to define in $I\Sigma_1$ a Δ_1 class ε linearly ordered by a Δ_1 ordering $\preceq$ with a least element 0 and with a Δ_1 operation Σ assigning to each finite non-empty decreasing sequence $\mu_1 \dots \mu_x$ of elements of ε and each sequence of non-zero numbers $a_1, \dots, a_x$ of the same length an element of ε denoted by $\sum_{i=1}^x \omega^{\mu_i} a_i$; the ordering is related to Σ as follows:

$$\sum_{i=1}^x \omega^{\mu_i} a_i \preceq \sum_{i=1}^y \omega^{\nu_i} b_i \text{ iff}$$

- (1) there is an $i \leq x, y$ such that $\mu_i \neq \nu_i$ or $a_i \neq b_i$, and, for the least such i , $\mu_i < \nu_i$ or $(\mu_i = \nu_i \text{ and } a_i < b_i)$ or
- (2) for each $i \leq x$, $\mu_i = \nu_i$ and $a_i = b_i$.

Furthermore, ε is least Δ_1 -class containing 0 and closed under Σ .

This is what we expect from ordinals $< \varepsilon$; we have to show that this can be achieved in $I\Sigma_1$. (We also expect well-order; but, as we shall see, this costs induction.) Thus let us make the following.

3.2 Definition. A *regular tree* is a set t of finite sequences of numbers such that

- (i) t contains with each s each initial segment of s , and
 - (ii) for each i, j, s if $s \frown \langle j \rangle \in t$ and $i < j$ then $s \frown \langle i \rangle \in t$.
- (Thus upper neighbours of s in t are $s \frown \langle 0 \rangle, s \frown \langle 1 \rangle, \dots, s \frown \langle i \rangle$ for some i).

A *pre-ordinal* is each regular tree t together with a mapping e (evaluation) assigning to each non-empty $s \in t$ a non-zero number $e(s)$. The *height* of t is the maximum of lengths of elements of t . We define an *operation* Σ applicable to each pair (μ, a) where μ is a non-empty sequence of pre-ordinals and a is a sequence of positive numbers such that $lh(\mu) = lh(a)$.

The pre-ordinal $(t, e) = \sum_1^x \omega^{\mu_i} a_i$ is defined as follows: let $\mu_i = (t_i, e_i)$ and let $t = \bigcup_{i=1}^x \{ \langle i \rangle \cap s \mid s \in t_i \} \cup \{ \emptyset \}$, $e(\langle i \rangle) = a_i$, and for $\emptyset \neq s \in t_i$ let $e(\langle i \rangle \cap s) = e_i(s)$. (It is easily seen that this corresponds to joining the evaluated trees $\mu_1, \dots, \mu_x$ over a new root and evaluating the old root of μ_i by a_i .)

Now define total Δ_1 functions $\mathcal{O}_{x,y}, \preccurlyeq_{x,y}$ as follows: $\mathcal{O}_{0,y} = \{ \emptyset \}$, $\mathcal{O}_{x+1,y} = \{ \sum_1^x \omega^{\mu_i} a_i \mid \mu_i \in \mathcal{O}_{x,y}, \mu_i \text{ descending } \preccurlyeq_{x,y}, a_i \leq y \}$; $\preccurlyeq_{x,y}$ using the obvious modifications of 3.1 (1), (2) above.

3.3 Fact. For each y , $\mathcal{O}_{x,y}$ is a total Δ_1 function of x ; $\mathcal{O}_{x,y} \subseteq \mathcal{O}_{x+1,y}$, $\mathcal{O}_{x,y} \subseteq \mathcal{O}_{x,y+1}$, analogously for $\preccurlyeq_{x,y}$. (Proofs in $I\Sigma_1$ evident).

3.4 Definition.

$$\begin{aligned} \mathcal{O}_y &= \bigcup_x \mathcal{O}_{x,y}, & \preccurlyeq_y &= \bigcup_x \preccurlyeq_{x,y} \\ \Omega_x &= \bigcup_y \mathcal{O}_{x,y}, & \preccurlyeq'_x &= \bigcup_y \preccurlyeq_{x,y} \\ \varepsilon &= \bigcup_{x,y} \mathcal{O}_{x,y}, & \preccurlyeq &\text{ is } \bigcup_{x,y} \preccurlyeq_{x,y} \end{aligned}$$

3.5 Fact. $\mathcal{O}_y, \preccurlyeq_y, \Omega_x, \preccurlyeq'_x, \varepsilon, \preccurlyeq$ are Δ_1 .

(Evidently, they are Σ_1 ; but for $\mu = (t, e)$ we have

$$\begin{aligned} \mu \in \mathcal{O}_y &\rightarrow \mu \in \mathcal{O}_{x,y} \text{ for } x = \text{height}(t) \\ \mu \in \Omega_x &\rightarrow \mu \in \mathcal{O}_{x,y} \text{ for } y = \max(\text{range}(e)) \end{aligned}$$

3.6 Fact. (1) $\preccurlyeq$ linearly orders ε . (2) ε is the smallest Δ_1 class X containing $\emptyset$ and closed under sums $\sum \omega^{\mu_i} a_i$ ($\mu_i \in X$ descending, a_i positive). (3) For each x , each non-empty Δ_1 subset of $\mathcal{O}_x$ has a $\preccurlyeq$ -least (i.e. $\preccurlyeq_x$ -least) element.

(To prove (2) show that each $\mathcal{O}_{x,y} \subseteq X$; to prove (3) observe that each $\mathcal{O}_{x,y}$ is a finite set and $\mathcal{O}_{x+1,y}$ is an end-extension of $\mathcal{O}_{x,y}$, i.e. each old element precedes each new one.)

3.7 Definition. 0 is *isolated*; $\sum_1^x \omega^{\mu_i} a_i$ is *isolated* if $\mu_x = 0$; otherwise it is a *limit ordinal* (or simply a *limit*).

3.8 Fact. μ is a limit iff $\mu > 0$ and has no predecessor.

3.9 Fact. Each $\mu \in \varepsilon$ has a successor.

3.10 Definition. For $\alpha = \sum_1^x \omega^{\mu_i} a_i$, $\beta = \sum_1^y \omega^{\nu_i} b_i$ define $\alpha \gg \beta$ iff $\nu_1 \preccurlyeq \mu_x$ (μ_x is the least exponent in α , ν_1 the greatest in β). Further we put $\alpha \gg 0$

if $\alpha \neq 0$. For $\alpha \gg \beta$ define $\alpha + \beta$ as follows: $\alpha + 0 = \alpha$; furthermore, if $\mu_x > \nu_1$ then $\alpha + \beta$ is given by exponents $\mu_1, \dots, \mu_x, \nu_1, \dots, \nu_y$ and coefficients $a_1, \dots, a_x, b_1, \dots, b_y$; if $\mu_x = \nu_1$ then it is given by exponents $\mu_1, \dots, \mu_x, \nu_2, \dots, \nu_y$ and coefficients $a_1, \dots, a_{x-1}, (a_x + b_1), b_2, \dots, b_y$.
(Thus e.g. $(\omega^3.3 + \omega^2.4) + (\omega^2.4 + \omega^0.7) = \omega^3.3 + \omega^2.8 + \omega^0.7$.)

3.11 Lemma. Each limit ordinal $\alpha \in \varepsilon$ can be uniquely written in the form $HD + \omega^\mu.1$ where $(HD \in \varepsilon \text{ (head) and } HD \gg \omega^\mu.1)$ or HD is empty and is disregarded. (Evident.)

3.12 Theorem. There is a Δ_1 function $\{\alpha\}(x)$ defined for each $\alpha \in \varepsilon$ and each x satisfying the following:

- (i) If $\alpha = \beta + 1$ then $\{\alpha\}(x) = \beta$; $\{0\}(x) = 0$;
if α is limit, $\alpha = HD + \omega^\mu.1$ ($\mu \geq 1$) then
$$\{\alpha\}(x) = \begin{cases} HD + \omega^{\mu-1} & \text{if } \mu \text{ is isolated } (\mu - 1 \text{ is the predecessor}); \\ HD + \omega^{\{\mu\}(x)} & \text{if } \mu \text{ is limit.} \end{cases}$$

For $x = 0$ the member $\omega^{\mu-1}x$ is deleted, thus the result is HD .

- (ii) $\alpha \neq 0$ implies $\{\alpha\}(x) < \alpha$;
 $\alpha \in \mathcal{O}_{x,y}$ and $z \leq y$ implies $\{\alpha\}(z) \in \mathcal{O}_{x,y}$;
 $\beta \gg \alpha > 0$ implies $\{\beta + \alpha\}(x) = \beta + \{\alpha\}(x)$ and $\beta \gg \{\alpha\}(x)$.

Proof. For each $y > 0$, define the function $D_y(\alpha, x) = \{\alpha\}_y(x)$ on $\mathcal{O}_y \times (\leq y)$ by the evident analogues of (i). Show that this is a Δ_1 function with domain $\mathcal{O}_y \times (\leq y)$ and that for $z > y$, D_y is a restriction of D_z . (To this end, show by induction on x that $D_y \upharpoonright (\mathcal{O}_{x,y} \times (\leq y))$ is a well-defined function with range included in $\mathcal{O}_{x,y}$. D_y is Σ_1 and its domain is Δ_1 ; thus D_y is Δ_1 .)

The last claim follows from the evident fact that, under given assumptions, $\beta + \alpha$ is isolated iff α is isolated and if they are limit then for $\alpha = HD + \omega^\mu$ we have $\beta + \alpha = (\beta + HD) + \omega^\mu$. □

3.13 Definition. $\alpha \xrightarrow{x} \beta$ means that there is a finite sequence $s = \langle \alpha_0, \dots, \alpha_r \rangle$ such that $\alpha_0 = \alpha$, $\alpha_r = \beta$ and, for $i < r$, $\alpha_{i+1} = \{\alpha_i\}(x)$. The sequence s is called the *witness* of $\alpha \xrightarrow{x} \beta$.

3.14 Lemma. (1) $\alpha \xrightarrow{x} \beta$ is Δ_1 . (2) If $X \subseteq \varepsilon$ is Σ_1 or Π_1 then so is $Y = \{\alpha \mid (\forall \beta)(\alpha \xrightarrow{x} \beta \rightarrow \beta \in X)\}$.

Proof. By induction on the length of the witness of $\alpha \xrightarrow{x} \beta$ show: if $\alpha \in \mathcal{O}_{x,y}$, $z \preccurlyeq x$ and $\alpha \xrightarrow{x} \beta$ then $\beta \in \mathcal{O}_{x,y}$. Let $F(\alpha, z) = \mathcal{O}_{\text{height}(\alpha), z}$ and $G(\alpha, z) =$ the set of all decreasing sequences of elements of $F(\alpha, z)$. F, G are Δ_1 and defined for all $\alpha \in \varepsilon$ and all z . The existential quantifier $(\exists s)$ in the above definition of $\alpha \xrightarrow{x} \beta$ may be replaced by $(\exists s \in G(\alpha, \max(e, z^*)))$ where z^* is

the maximal number from all the evaluations in α and z . This shows that $\alpha \xrightarrow{z} \beta$ is Δ_1 . The proof of (2) is similar. $\square$

3.15 Theorem (Properties of $\rightarrow$).

- (1) $\alpha \xrightarrow{z} \beta \xrightarrow{z} \gamma$ implies $\alpha \xrightarrow{z} \gamma$
- (2) If $\beta \gg \alpha > 0$ and $\alpha \xrightarrow{z} \gamma$ then $\beta + \alpha \xrightarrow{z} \beta + \gamma$
- (3) $\alpha \xrightarrow{z} 0$; more generally, if $\alpha \geq \omega^0.k$ and $z \geq k$ then $\alpha \xrightarrow{z} \omega^0.k$
- (4) $x < y$ implies $\omega^\alpha.y \xrightarrow{z} \omega^\alpha.x$
- (5) $\omega^{\delta+1} \xrightarrow{z} \omega^\delta$ for each $z > 0$
- (6) $\alpha \xrightarrow{z} \beta$ implies $\omega^\alpha \xrightarrow{z} \omega^\beta$
- (7) $x < y$ implies $\{\alpha\}(y) \xrightarrow{1} \{\alpha\}(x)$
- (8) $x < y$ and $\alpha \xrightarrow{x} \beta$ implies $\alpha \xrightarrow{y} \beta$
- (9) $x \geq 1$, $\alpha \xrightarrow{x} \beta$ and $\alpha > \beta$ implies $\alpha \xrightarrow{x+1} \beta + 1$
- (10) for $\alpha > 0$ and $x > 1$, $\omega^\alpha \xrightarrow{x} \omega^{\{\alpha\}(x-1).x}$
- (11) $\alpha \xrightarrow{x} \beta$ implies $\{\alpha\}(x) \xrightarrow{x} \{\beta\}(x)$;
if $\alpha > \beta$ and $\alpha \xrightarrow{x} \beta$ then $\{\alpha\}(x) \xrightarrow{x} \beta$.

Remark. Needless to say, ω^α stands for $\omega^\alpha.1$, i.e. for the corresponding one-element sum.

Proofs. (2) follows from the last claim in 3.12 (ii). (Inspect, by induction, each number of the witnessing sequence.)

(3) Let $k \leq z \leq x$. Let α be the smallest element of $\mathcal{O}_x$ such that $\alpha \geq \omega^0.k$ and not $\alpha \xrightarrow{z} \omega^0.k$, then $\alpha > \omega^0.k$, $\{\alpha\}(z) < \alpha$, $\{\alpha\}(z) \in \mathcal{O}_x$ and one shows by checking the defining properties in 3.12 that $\{\alpha\}(z) \geq \omega^0.k$; thus $\{\alpha\}(z) \xrightarrow{z} \omega^0.k$.

(4) By (3), $\omega^\alpha(y-x) \xrightarrow{z} 0$; furthermore, $\omega^\alpha.x \gg \omega^\alpha.(y-x)$. Thus, by (2), $\omega^\alpha.y \xrightarrow{z} \omega^\alpha.x$.

(5) $\omega^{\delta+1} \xrightarrow{z} \omega^\delta.z \xrightarrow{z} \omega^\delta$ (by (4)).

(6) It is enough to show: if $\beta = \{\alpha\}(x)$ then $\omega^\alpha \xrightarrow{z} \omega^\beta$. First assume α a limit: then $\{\omega^\alpha\}(z) = \omega^{\{\alpha\}(z)} = \omega^\beta$. Now let $\alpha = \gamma + 1$, then $\beta = \gamma$ and, by (5), $\omega^\alpha \xrightarrow{z} \omega^\beta$.

(7) Trivial for α isolated; assume α limit. Let $x, y \leq z$ and $\alpha \in \mathcal{O}_z$; we proceed by induction in $\mathcal{O}_z$. Let $\alpha = HD + \omega^\mu$; then $\omega^\mu \in \mathcal{O}_z$. If HD is non-empty, then the induction assumption gives $\{\omega^\mu\}(y) \xrightarrow{1} \{\omega^\mu\}(x)$ and the result follows by (2). Thus assume HD empty, $\alpha = \omega^\mu$. If μ is isolated

then see (4); if it is a limit then $\{\omega^\mu\}(y) = \omega^{\{\mu\}(y)} \xrightarrow{1} \omega^{\{\mu\}(x)} = \{\omega^\mu\}(x)$ by the induction assumption (since $\mu, \{\mu\}(y), \{\mu\}(x) \in \mathcal{O}_z$).

(8) Assume $\alpha \in \mathcal{O}_z$, $z > x, y$, $\beta = \{\alpha\}(x)$ and use induction on α : $\alpha \xrightarrow{y} \{\alpha\}(y) \xrightarrow{1} \{\alpha\}(x)$ by (7), thus $\{\alpha\}(y) \xrightarrow{y} \{\alpha\}(x) = \beta$ and $\alpha \xrightarrow{y} \beta$.

(9) It suffices to assume $\{\alpha\}(x) = \beta$ and to prove $\alpha \xrightarrow{x+1} \beta + 1$. To this end, it suffices to prove the following for each fixed z and $x + 1 \leq z$, $x \geq 1$:

For each $\alpha \in \mathcal{O}_z$, if $\{\alpha\}(x) = \beta$ then $(\beta + 1) \in \mathcal{O}_z$ and $\alpha \xrightarrow{x+1} \beta + 1$. This is proved by induction on α . The case of α being isolated is trivial; assume $\alpha = HD + \omega^\delta$. First assume HD non-empty and put $\{\omega^\delta\}(x) = \beta_0$. Then $\{\alpha\}(x) = HD + \beta_0$ and, by the induction assumption, $\omega^\delta \xrightarrow{x+1} \beta_0 + 1$ and $\beta_0 + 1 \in \mathcal{O}_z$. Thus $\alpha = HD + \omega^\delta \xrightarrow{x+1} HD + \beta_0 + 1 = \{\alpha\}(x) + 1 = \beta + 1$ and one easily sees that $\beta + 1 \in \mathcal{O}_z$ ($\beta_0 + 1 \in \mathcal{O}_z$ and the last exponent in HD is strictly greater than the exponent of β_0).

Assume HD empty, thus $\alpha = \omega^\delta$. If δ is $\mu + 1$ then $\alpha \xrightarrow{x+1} \{\omega^\delta\}(x + 1) = \omega^\mu(x + 1) = \omega^\mu \cdot x + \omega^\mu \xrightarrow{x+1} \omega^\mu \cdot x + 1$ (since, by (3), $\omega^\mu \xrightarrow{x+1} 1$), thus $\alpha \xrightarrow{x+1} \beta + 1$. If δ is a limit then $\omega^\delta \xrightarrow{x+1} \omega^{\{\delta\}(x+1)} \in \mathcal{O}_z$ (since $x + 1 \leq z$); by (7) and (6) $\omega^{\{\delta\}(x+1)} \xrightarrow{1} \omega^{\{\delta\}(x)}$ and $\xrightarrow{1}$ may be replaced by $\xrightarrow{x}$ (by (8) since $1 \leq x$). By the induction hypothesis we get $\omega^{\{\delta\}(x+1)} \xrightarrow{x+1} \omega^{\{\delta\}(x)} + 1$ which gives $\alpha = \omega^\delta \xrightarrow{x+1} \omega^{\{\delta\}(x)} + 1 = \beta + 1$ as desired.

(10) Assuming $\alpha > 0$ and $x > 1$ we prove $\omega^\alpha \xrightarrow{x} \omega^{\{\alpha\}(x-1)} \cdot x$. Clearly, $\alpha \xrightarrow{x-1} \{\alpha\}(x-1)$ (and $x-1 \geq 1$); by (9), $\alpha \xrightarrow{x} \{\alpha\}(x-1) + 1$. Thus, by (6), $\omega^\alpha \xrightarrow{x} \omega^{\{\alpha\}(x-1)+1} \xrightarrow{x} \omega^{\{\alpha\}(x-1)} \cdot x$ and we are done.

(11) This is a triviality: $\alpha = \beta$ implies $\{\alpha\}(x) = \{\beta\}(x)$ and $\alpha > \beta$ and $\alpha \xrightarrow{x} \beta$ implies that the witnessing sequence is $\alpha, \{\alpha\}(x), \dots, \beta$, thus $\{\alpha\}(x) \xrightarrow{x} \beta$. This gives $\{\alpha\}(x) \xrightarrow{x} \{\beta\}(x)$. $\square$

3.16 Definition. For each $\mu \in \varepsilon$ we put $\omega_0^\mu = \mu$, $\omega_{x+1}^\mu = \omega_x^\mu$.

3.17 Remark.

(1) Evidently, if $y \geq 1$ and $\mu \in \mathcal{O}_y$ then $\omega_x^\mu \in \mathcal{O}_y$ for each x .

(2) If $\mu = \omega_0^n$ we shall write ω_x^n instead of ω_x^μ .

(3) Evidently,

$$\omega_x^y \xrightarrow{z} \omega_x^{y-1} \quad \text{for } y \geq 1 \text{ and } z > 0;$$

furthermore, for $x > 0$, and $z > 0$,

$$\omega_x^1 \xrightarrow{z} \omega_{x-1}^z$$

(note that $\omega_x^0 = \omega_{x-1}^1$ for $x > 0$). The first relation is by 3.15 (5,6), the second from $\omega^1.1 \xrightarrow{z} \omega^0.z$ by (6).

(4) Also evidently, for each $\alpha \in \varepsilon$, and each x , $\alpha \in \Omega_x$ iff $\alpha < \omega_x^1$. (Prove this Π_1 -property of x by induction on x .)

(b) Transfinite Induction and Fragments

3.18 Theorem. For each $m, k, n \geq 1$, $I\Sigma_{m+k-1}$ proves the following: each non-empty Σ_m set of ordinals less than ω_k^n has a least element in the ordering $\preceq$.

Proof. By induction on k . First, for each n , $I\Sigma_m$ proves that each non-empty Σ_m subset of $(\preceq \omega_1^n)$ has a least element: this can be proved by induction on n . The case $n = 1$ is clear since $I\Sigma_1$ proves that $(< \omega_1^1)$ is Δ_1 isomorphic with the universe of all numbers with $\leq$. It also proves that ω_1^1 is Δ_1 isomorphic to the cartesian power of n copies of the universe ordered lexicographically. Assume we have proved on $I\Sigma_m$ the claim for n and let $X \subseteq \omega \times \cdots \times \omega$ $(n+1)$ times.

Let a be the least number such that, for some n -tuple s , $(\langle a \rangle \frown s) \in X$. Such an a exists since the condition is Σ_m . Let Y be the set of all such sequences s ; by the induction assumption, we can prove that Y has a least element s_0 in the lexicographic ordering. Thus $\langle a \rangle \frown s_0$ is least in X ; this completes the proof in $I\Sigma_m$.

Now assume we can prove the theorem for k, n and $(m+1)$; we show that it holds for $(k+1), n$ and m . This will complete the whole proof. We proceed in $I\Sigma_{k+m}$. Let $X \neq \emptyset$ be Σ_m , $X \subseteq (\leq \omega_{k+1}^n)$. Define a function F as follows:

$F(0)$ is the minimal (μ_0, a_0) such that $\mu_0 \in (< \omega_k^n)$, $a_0 > 0$ and X contains an element $\omega^{\mu_0} a_0 + \dots$. (Existence clear; minimality is understood lexicographically, using $<$ and $\leq$.) Let $F(x) = (\mu_x, a_x)$ be given; we define $F(x+1)$.

$F(x+1)$ is the minimal (μ_{x+1}, a_{x+1}) such that $\mu_{x+1} \in (< \omega_k^n)$, $a_{x+1} > 0$ and X contains an element $\omega^{\mu_0} a_0 + \dots + \omega^{\mu_x} a_x + \omega^{\mu_{x+1}} a_{x+1} + \dots$ if there is such an element; otherwise $F(x+1) = F(x)$.

Clearly, such F is well-defined in $I\Sigma_{k+m}$ (it is Δ_{m+1}) and the set $Y = \{\mu_x \mid x\}$ is Σ_{m+1} and non-empty; furthermore, $Y \subseteq (\preceq \omega_k^n)$. By the induction hypothesis Y has a least element μ ; μ is μ_y for some y . But this means that the element $\sum_1^y \omega^{\mu_x} a_x$ is the least element of X . $\square$

Remark. Let $L(\omega_k^n, \Sigma_m)$ be the statement "each non-empty Σ_m subset of $(< \omega_k^n)$ has a least element". Thus we proved, for each $m, k, n \geq 1$,

$$I\Sigma_{m+k-1} \vdash L(\omega_k^n, \Sigma_m).$$

(c) α -large Sets in $I\Sigma_1$

The notion of an α -large set is technically very useful and is also appealing in its own right. In the next subsection (d) we shall show that it is naturally related to the Schwichtenberg-Wainer hierarchy of functions.

3.19 Definition. Let A be a finite set and let $(a_0, \dots, a_q)$ be its increasing enumeration; we define $\{\alpha\}A$ for each $\alpha \in \varepsilon$. If $A = \emptyset$ then $\{\alpha\}A = \alpha$; otherwise we put

$$\{\alpha\}(a_0, \dots, a_q) = \{\{\alpha\}(a_0)\}(a_1, \dots, a_q).$$

(Clearly, this defines $\{\alpha\}A$ as a total Δ_1 function.) A is α -large if $\{\alpha\}A = 0$.

3.20 Remark. (1) A is x -large (i.e. ω^0 - x -large) if and only if $\text{card}(A) \geq x$.

(2) A is ω -large (i.e. ω^1 -1-large) iff $\text{card}(A) > \min A$.

(3) $A = (a_0, \dots, a_q)$ is α -large iff $A - (a_0, \dots, a_i)$ is $\{\alpha\}(a_0, \dots, a_i)$ -large. (Evident.)

3.21 Theorem (1) If A is α -large, $x \leq \min A$ and $\alpha \xrightarrow{x} \beta$ then A is β -large.

(2) If $A = (a_0, \dots, a_q)$, $B = (b_0, \dots, b_r)$, $q \leq r$, for $i \leq q$ we have $b_i \leq a_i$ and A is α -large then B is β -large; (in particular) if $A \subseteq B$ and A is α -large then B is α -large.

(3) Let $\alpha \gg \beta > 0$. Then A is $(\alpha + \beta)$ -large iff there are B, C such that $A = B \cup C$, $\max B < \min C$, B is β -large and C is α -large.

(4) Let $\alpha \geq 1$, $A = (a_0, \dots, a_q)$, $a_0 \geq 2$, $a_0 = x_0 < \dots < x_{a_0} = a_q$ (i.e. $[a_0, a_q]$ is decomposed into a_0 intervals). If A is ω^α -large, then there is an $i < a_0$ such that $(x_i, x_{i+1}] \cap A$ is $\omega^{\{\alpha\}(a_0-1)}$ -large.

(Here

$$(x_i, x_{i+1}] = \{z \mid x_i < z \leq x_{i+1}\}.)$$

(5) If $A = (a_0, \dots, a_q)$ is ω_x^y -large, and $a_0, y \geq 1$ then both $(a, \dots, a_q)$ and $(a_0, \dots, a_{q-1})$ are ω_x^{y-1} -large.

Proofs. (1) By 3.15 (8), we may assume $x = \min A$. Put $\alpha_i = \{\alpha\}(a_0, \dots, a_{i-1})$, $\beta_i = \{\beta\}(a_0, \dots, a_{i-1})$; thus $\alpha_0 = \alpha$, $\beta_0 = \beta$, $\alpha_{q+1} = 0$, $\alpha_0 \xrightarrow{a_0} \beta_0$.

By 3.15 (11), $\alpha_1 \xrightarrow{a_0} \beta_1$ thus $\alpha_1 \xrightarrow{a_1} \beta_1$; similarly we get $\alpha_i \xrightarrow{a_i} \beta_i$, thus $\alpha_{q+1} \xrightarrow{a_q} \beta_{q+1}$, i.e. $0 = \alpha_{q+1} \geq \beta_{q+1} = 0$.

(2) Let $A = (a_0, \dots, a_q)$, $B = (b_0, \dots, b_r)$, $r \geq q$; we have $b_i \leq a_i$ for $i \leq q$. Let $\alpha_i = \{\alpha\}(a_0, \dots, a_{i-1})$, $\beta_i = \{\alpha\}(b_0, \dots, b_{i-1})$, thus $\alpha_{i+1} = \{\alpha_i\}(a_i)$ and similarly for β . First, $\alpha_1 = \{\alpha\}(a_0) \xrightarrow{1} \{\alpha\}(b_0) = \beta_1$ (for $a_0 \geq 1$ by 3.15 (7)), thus $\alpha_1 \xrightarrow{a_0} \beta_1$ by 3.15 (8). Now assume

$\alpha_{i+1} \xrightarrow{a_i} \beta_{i+1}$, $i + 1 \leq q$. If $\beta_{i+1} = 0$ then $\{\alpha\}(B) = 0$; if $\beta_{i+1} > 0$ then $\alpha_{i+1} \xrightarrow{a_{i+1}} \beta_{i+1} \xrightarrow{a_{i+1}} \{\beta_{i+1}\}(a_{i+1}) \xrightarrow{a_{i+1}} \{\beta_{i+1}\}(b_{i+1}) = \beta_{i+2}$ and $\alpha_{i+1} > \beta_{i+2}$; by 3.15 (11) $\alpha_{i+2} = \{\alpha_{i+1}\}(a_{i+1}) \xrightarrow{a_{i+1}} \beta_{i+2}$. Thus in any case $\{\alpha\}B = 0$.

(3) Let $\alpha, \beta > 0$, $\alpha \gg \beta$. First assume $A = (a_0, \dots, a_q)$ to be $(\alpha + \beta)$ -large. Put $\lambda_i = \{\alpha + \beta\}(a_0, \dots, a_{i-1})$, $\beta_i = \{\beta\}(a_0, \dots, a_{i-1})$. We have $\lambda_{q+1} = 0$; thus, for some i , $\beta_i = 0$. Let m be minimal such i . Then $\lambda_m = \alpha$, $(a_0, \dots, a_{m-1}) = B$ is β -large and $(a_m, \dots, a_q) = C$ is α -large.

Conversely, let $A = B \cup C$, $\max B < \min C$, let B be β -large, C α -large; assume that B is the least possible. Then $\{\beta\}(B) = 0$, $\{\alpha + \beta\}(B) = \alpha$, $\{\alpha + \beta\}(B \cup C) = \{\alpha\}(C) = 0$.

(4) By 3.15 (10), $\omega^\alpha \xrightarrow{a_0} \omega^{\{\alpha\}(a_0-1)} \cdot a_0$; thus by 3.15 (11), $\{\omega^\alpha\}(a_0) \xrightarrow{a_0} \omega^{\{\alpha\}(a_0-1)} \cdot a_0$. If $A = (a_0, \dots, a_q)$ is ω^α -large then $A - (a_0)$ is $\{\omega^\alpha\}(a_0)$ -large and therefore $\omega^{\{\alpha\}(a_0-1)} a_0$ -large. By (3) this means that $A - (a_0)$ may be decomposed into $B_1, \dots, B_a$ that are mutually disjoint and such that $\max B_i < \min B_{i+1}$ and each B_i is $\omega^{\{\alpha\}(a_0-1)}$ -large. Thus if we have the decomposition $a_0 = x_0 < \dots < x_{a_0} = a_q$ of $[a_0, a_q]$, at least one half-closed interval $(x_i, x_{i+1}]$ must contain B_i and therefore is $\omega^{\{\alpha\}(a_0-1)}$ -large.

(5) Evidently, $(a_1, \dots, a_q)$ is $\{\omega_x^y\}(a_0)$ -large; since $\omega_x^y \xrightarrow{a_0} \omega_x^{y-1}$ we have $\{\omega_x^y\}(a_0) \xrightarrow{a_0} \omega_x^{y-1}$ (3.17); the result follows by (1). For $(a_0, \dots, a_q)$ use (2). □

(d) Schwichtenberg-Wainer Hierarchy

3.22 We shall investigate the hierarchy of number theoretic functions defined informally as follows:

$$\begin{aligned} f_0(x) &= x + 1 \\ f_{\alpha+1}(x) &= f_\alpha^x(x + 1) \\ f_\lambda(x) &= f_{\{\lambda\}(x)}(x + 1) \text{ for } \lambda \text{ limit.} \end{aligned}$$

Here α, λ vary over ordinals $< \varepsilon$; $f^x(y)$ means x -th iteration of f , i.e. $f^0(y) = y$, $f^{x+1}(y) = f(f^x(y))$. This is a variant of the hierarchy investigated by Schwichtenberg and Wainer and the second from two hierarchies investigated by Solovay and Ketonen. We shall show that this hierarchy is Δ_1 definable in $I\Sigma_1$ as a hierarchy of partial functions, show conditions sufficient to prove that a given function f_α is total and relate the hierarchy to α -large sets (cf. Theorem 3.30).

3.23 Remark. (1) It is easy to show in $I\Sigma_1$ that if F is a total one-argument Δ_1 function then there is a unique total two-argument function G such that $G(x, y) = F^x(y)$ for each x, y .

(2) Similarly, $I\Sigma_1$ proves that if q is a finite one-argument function then there is a finite two-argument function q' which is the maximal function such that for each $(x, y) \in \text{dom}(q')$, $q'(x, y) = q^x(y)$. Easy proofs are left to the reader.

(3) We shall work with finite two-argument functions defined for *some* pairs (α, x) where $\alpha \in \varepsilon$ and x is a number. If q is such a function then q_α will be the unique function such that for all x , $q_\alpha(x)$ is defined iff $q(\alpha, x)$ is defined and then $q_\alpha(x) = q(\alpha, x)$. (Needless to say, " $q_\alpha(x)$ is defined" means $x \in \text{dom } q_\alpha$.)

3.24 Definition. Define a predicate $WD(q, \alpha, x, y)$ (read " q is a derivation of $f(x) = y$ " or, pedantically, " q is a derivation of the fact that the value of x in the α -th function in the Schwichtenberg-Wainer hierarchy is y ") as follows:

- (1) q is a finite function, $\text{dom}(q) \subseteq \varepsilon \times \omega$;
- (2) $q_\alpha(x) = y$
- (3) wherever $q_\beta(z)$ is defined then
 - (i) if $z > 0$ then $q_\beta(z-1)$ is defined,
 - (ii) if $\beta = \gamma + 1$ then $q_\gamma^z(z+1)$ is defined and equal to $q_\beta(z)$,
 - (iii) if β is a limit then $q_{\{\beta\}(z)}(z+1)$ is defined and equal to $q_\beta(z)$.

3.25 Lemma. (1) WD is Δ_1 . (2) If $WD(q, \alpha, x, y)$, $q_\beta(z)$ is defined and $\beta \xrightarrow{z} \gamma$ then $q_\gamma(z)$ is defined. (3) $WD(q, \alpha, x, y)$ and $WD(q', \alpha, x, y')$ implies $y = y'$.

Proof (in $I\Sigma_1$) is easy. (1) Note that ε is Δ_1 ; all "is defined"-quantifiers can be bounded by q . It remains to observe that " $u = f^x(y)$ " is Δ_1 in f, u, x, y .

(2) Assume $\gamma = \{\beta\}(z)$. If β is limit then use (iii) and (i); if β is $\beta_0 + 1$ then $\{\beta\}(z) = \beta_0$ and use (ii) and (i). $\square$

3.26 Lemma. Assume $WD(q, \alpha, x, y)$ and $q_\beta(z)$ defined. Then

- (1) $q_\beta(z) \geq z + 1$
- (2) $w < z$ implies $q_\beta(w) < q_\beta(z)$
- (3) $z > 0$, $\beta \xrightarrow{z} \gamma$ and $\beta \neq \gamma$ implies $q_\beta(z) > q_\gamma(z)$.

Proof. Prove simultaneously (1) & (2) & (3) by induction on β running over all γ such that $q_\gamma \neq \emptyset$. $\square$

3.27 Corollary. Let $WD(q, \alpha, x, y)$, $\alpha \in \mathcal{O}_{u,v}$ and $y \leq v$. Then the restriction q' of q to $\mathcal{O}_{u,v} \times (\leq y)$ satisfies $WD(q', \alpha, x, y)$.

Proof. Verify the conditions 3.23 (i), (ii), (iii) for q' by induction on $\beta \in \mathcal{O}_{u,v}$ using 3.26 (1), (2). $\square$

3.28 Definition. $y = f_\alpha(x)$ iff $(\exists q) WD(q, \alpha, x, y)$.

$$[x, y] = \{z \mid x \leq z < y\}; \quad [(x, y)] = \{z \mid x < z < y\}; \quad \text{similarly } [(x, y)].$$

3.29 Remark. (1) Clearly this defines a partial two-argument function. Observe that 3.27 implies that this function is Δ_1 .

Note also that by 3.26, the functions satisfy the following whenever defined:

- (i) $f_\alpha(z) \geq z + 1$,
- (ii) $w < z$ implies $f_\alpha(w) < f_\alpha(z)$,
- (iii) $z > 0$, $\beta \xrightarrow{z} \gamma \neq \beta$ implies $f_\beta(z) > f_\gamma(z)$.

(2) Recall that, for $k \geq 1$ and any n , $I\Sigma_{k+1} \vdash L(\omega_k^n, \Sigma_2)$. Using this we may prove in $I\Sigma_{k+1}$ that for each limit $\gamma \prec \omega_k^n$, $\gamma = \sup_x \{\gamma\}(x)$. (This is a Π_2 condition on γ ; for its proof, $L(\omega_k^n, \Sigma_2)$ is sufficient: consider the least γ not satisfying this.) Also observe that $I\Sigma_{k+1}$ proves that for each $\alpha \prec \beta \prec \omega_k^n$ there is a z such that $\beta \xrightarrow{z} \alpha$. (This again is Π_2 in β .)

3.30 Theorem. For $k \geq 0$ and any n , $I\Sigma_{k+1}$ proves the following:

- (1) For each $\alpha \prec \omega_k^n$, f_α is total.
- (2) For each $\alpha \prec \beta \prec \omega_k^n$, there is a z such that

$$(\forall w > z)(f_\alpha(w) < f_\beta(w)).$$

- (3) For each $\alpha \prec \omega_k^n$ and each x , $f_\alpha(x)$ is the least y such that the interval $[x, y]$ is ω^α -large.

Proof. (1) For $k = 0$ this follows by applying n times Remark 3.23. For $k > 0$ observe that the statement in question is Π_2 so that $L(\omega_k^n, \Sigma_2)$ suffices to prove it for all $\alpha \prec \omega_k^n$ (using Remark 3.23).

(2) By 3.29 (2), take a z such that $\beta \xrightarrow{z} \alpha$; by 3.29 (1) (iii), this implies $f_\alpha(z) < f_\beta(z)$. If $w > z$ then, by 3.15 (8), we have $\beta \xrightarrow{w} \alpha$ and therefore $f_\alpha(w) < f_\beta(w)$.

(3) We shall proceed by induction on $\alpha \prec \omega_k^n$; observe that assuming totality of all f_α , $\alpha \prec \omega_k^n$, the assertion in question is Π_1 . The case $\alpha = 0$ is clear.

Claim. Assume the assertion of (3) for α ; then for each $y \geq 1$ and each x , $[x, f_\alpha^y(x)]$ is the minimal $(\omega^\alpha \cdot y)$ -large interval beginning with x . Proof by induction on y (the present assertion is Δ_1 in y). For $y = 1$ this is our assumption; assume the present assumption for $y-1$. Then $[x, z]$ is $\omega^\alpha \cdot y$ -large iff $[f_\alpha(x), z]$ is $\omega^\alpha \cdot (y-1)$ -large (by 3.21 (3)) iff $z \geq f_\alpha^{y-1}(f_\alpha(x)) = f_\alpha^y(x)$. This proves the claim.

Continuing the proof of (3), consider $\alpha + 1$. For $x = 0$ we easily see that $f_{\alpha+1}(0) = 1$, and $\{\omega^{\alpha+1}\}(0) = 0$, thus the one-element set $\{0\}$ is $\omega^{\alpha+1}$ -large. Thus assume $x > 0$ and use the claim: $[x, z]$ is $\omega^{\alpha+1}$ -large iff $[x+1, z]$ is $\omega^\alpha \cdot x$ -large, iff $z \geq f_\alpha^x(x+1) = f_{\alpha+1}(x)$.

It remains to consider α being limit. Then $[x, z]$ is ω^α -large iff $[x+1, z]$ is $\omega^{\{\alpha\}(x)}$ -large iff $z \geq f_{\{\alpha\}(x)}(x+1) = f_\alpha(x)$. This completes the proof. $\square$

3.31 Remark. The reader may verify the following as an exercise:

$$\begin{aligned} f_1^y(x) &= 2^y(x+1) - 1 \\ f_2(x) &= 2^x(x+2) - 1. \end{aligned}$$

3.32 Theorem. $I\Sigma_1$ proves the following: for each α, x, z , $z = f_\alpha(x)$ iff z is minimal such that $[x, z]$ is ω^α -large.

(Observe that we do not claim that $f_\alpha(x)$ exists, but we claim that if it exists and equals z then $[x, z]$ is ω^α -large and z is minimal with that property; and if there is a z such that $[x, z]$ is ω^α -large and z is minimal with this property then $f_\alpha(x)$ exists and equals z . Our proof is an inspection of the proof of 3.30 (2).)

Proof. Let $\alpha \in \mathcal{O}_q$ and $x, z \leq q \geq 1$. We prove by induction on $\alpha \in \mathcal{O}_q$ the following Δ_1 property of α :

(*) $(\forall z \leq q)(\forall x \leq q)(z = f_\alpha(x) \text{ iff } z \text{ is minimal such that } [x, z] \text{ is } \omega^\alpha\text{-large})$.

This is clear for $\alpha = 0$. Assume (*) for α and let $\alpha + 1 \in \mathcal{O}_y$.

Claim. For all $1 \leq y \leq q$, $z \leq q$, $z = f_\alpha^y(x)$ iff z is minimal such that $[x, z]$ is $(\omega^\alpha \cdot y)$ -large. (See the proof of 3.30.)

We may assume $x \geq 1$. By the claim, $[x, z]$ is $\omega^{\alpha+1}$ -large iff $[x+1, z]$ is $\omega^\alpha \cdot x$ -large iff $z \geq f_\alpha^x(x+1) = f_{\alpha+1}(x)$ – as in 3.30. Similarly for α being limit. $\square$

3.33 Definition. Let $(W)_u$ be the formula

$$(\forall x, z)(\exists y)([x, y] \text{ is } \omega_u^z\text{-large})$$

(the principle of ordinal-large intervals).

3.34 Facts. (1) $(W)_u$ is a Π_2 -formula.

(2) $I\Sigma_1 \vdash (\forall u)((W)_u \equiv (\forall z)(\forall \alpha < \omega_{u-1}^z)(f_\alpha \text{ is total}))$

(3) For each $k \geq 0$, $I\Sigma_{k+1} \vdash (W)_k$ and, for each n , $I\Sigma_{k+1} \vdash (\forall x)(\exists y)([x, y] \text{ is } \omega_{k+1}^n\text{-large})$.

((2) follows by Theorem 3.32.)

3.35 Corollary. $I\Sigma_1$ proves the following:

$$(\forall u)(\text{Con}(I\Sigma_u + \text{Tr}(\Pi_1)) \rightarrow (W)_u).$$

This follows from 3.30 (formalized in $I\Sigma_1$) exactly as the analogous statement in 2.17.

3.36 Theorem. For each k , $I\Sigma_1$ proves $(PH)_k \equiv (W)_k$; thus it proves $(W)_k \equiv \text{Con}(I\Sigma_k + \text{Tr}(\Pi_1))$.

Comment. One implication (easy) is 3.35. For the converse, thanks to the main result of Sect. 2 is enough to show in $I\Sigma_1$ the following:

$$(*) \quad (\forall u)((W)_u \rightarrow (PH)_u),$$

or, at least, to prove each instance of this.

Here we have two possibilities:

(a) Solovay and Ketonen proved that, for each $k \geq 1$, $c \geq 2$, $b > a \geq 3$, if $[a, b]$ is ω_k^{c+5} -large then $[a, b] \rightarrow_* (k+2)_c^{k+1}$. If one checks that their proof works in $I\Sigma_1$ (which we expect but have not checked) then the implication (*) is proved.

(b) Paris has a model-theoretic proof of $(W)_k \rightarrow (PH)_k$ (for any standard k). We shall elaborate it in Chap. IV (see IV.3.37).

3.37 Problem. Find a reasonably simple proof of $I\Sigma_1 \vdash (W)_u \rightarrow (PH)_u$ or, at least, $I\Sigma_1 \vdash (\forall u)(W)_u \rightarrow (\forall u)(PH)_u$ or, alternatively, $I\Sigma_1 \vdash (\forall u)(W)_u \rightarrow \text{Con}(PA + \text{Tr}(\Pi_1))$. Are details of Solovay-Ketonen's paper dispensable?

For we know in part, and we prophesy in part.
But when that which is perfect is come, then
that which is in part shall be done away.
(I. Corinthians 13, 9–10)

Part B

Incompleteness

Chapter III

Self-Reference

Introduction. The preceding Part A was devoted to positive results in fragments of arithmetic: recall that we already showed that Robinson's Q proves all true Σ_1 sentences (which was rather easy) and then devoted considerable space to the development of mathematics in fragments ($I\Sigma_n$ or $B\Sigma_{n+1}$). We now know how to develop a theory of finite sets and definable infinite sets in $I\Sigma_1$ and we have arithmetized important parts of logic, recursion theory and combinatorics, sometimes using fragments stronger than $I\Sigma_1$. The present Part B will deal with *incompleteness* of systems of axioms in arithmetic: we shall prove the celebrated Gödel's incompleteness theorems (saying among other things that no consistent axiomatized theory containing Q is complete, i.e. each such theory has a sentence φ such that $(T + \varphi)$ and $(T + \neg\varphi)$ is consistent. Moreover, we shall be interested in comparing theories (containing arithmetic) with respect to their strength. One possibility is just to investigate the inclusion of theories (T is a subtheory of S , i.e. each formula provable in T is provable in S), but we shall study two related notions, namely interpretability (one can define basic notions of T in S such that T becomes a subtheory of S modulo these definitions) and partial conservativity (for some class Γ of formulas, e.g. Σ_k -formulas, each $\varphi \in \Gamma$ provable in T is provable in S). The study will be confined to theories containing a certain fragment of arithmetic.

The present Chapter III is devoted to Gödel's method of self-reference (which is of proof-theoretical character); the next Chapter IV will deal with models of fragments. Note that self-reference has interesting philosophical aspects; but they will be entirely disregarded. Chapter III has the following structure:

Section 1 contains preliminaries, in particular it presents a definition of a theory *containing* (some) *arithmetic*. This is important since the results of this chapter are not confined to theories in the language of arithmetic (and apply, for example, also to systems of set theory). Section 2 contains Gödel's incompleteness theorems and related topics, as well as a characterization of interpretability among theories like PA (having induction for all formulas). In Sect. 3 we shall deal with theories not having induction for all

formulas, mainly with finitely axiomatized theories; we present a strengthening of Gödel's second incompleteness theorem and a characterization of interpretability for such theories. Finally in Sect. 4 we shall systematically study interpretability and partial conservativity with a special emphasis to finitely axiomatizable theories. Among other theories we shall study the system ACA_0 – finitely axiomatizable second order arithmetic extending PA conservatively – and show how these theories differ with respect to interpretability. (This will be obtained as a corollary to general theorems on interpretability and partial conservativity.)

1. Preliminaries

(a) Interpretability and Partial Conservativity

This subsection may be viewed as a continuation of Sect. 0 (Preliminaries) at the beginning of the book. We collect here several definitions and state some obvious facts. Recall the notion of a theory T (0.11) and its language L . It is clear what we mean by saying that T and L are Δ_1 .

1.1 Definition. (1) T is a *subtheory* of S (or: S is an *extension* of T) if the language L_T of T is a sublanguage of the language L_S of S and if each formula of L_T provable in T is provable in S .

(2) Let Γ be a class of formulas. T is Γ -*conservative* over S if each Γ -formula φ of L provable in T is provable in S .

(3) S is a Γ -*conservative extension* of T if S is an extension of T , i.e. T is a subtheory of S , and, in addition, S is Γ -conservative over T ; i.e. if *any* formula is T -provable then it is S -provable; and if a Γ -formula is S -provable then it is T -provable. (Thus S is stronger than T but for Γ -formulas T -provability coincides with S -provability.)

1.2 Definition. (1) Let L_1 be a language and let S be a theory in a language L_2 . To *interpret* L_1 in S means to define the following in S :

- the range of variables of L_1 -formulas in S and
- for each predicate P , each function symbol F and each constant c of L_1 , its translation P^* , F^* , c^* into S .

In more details, such an interpretation $*$ is given by:

- a formula $\chi(x)$ such that $S \vdash (\exists x)\chi(x)$,
- for each n -ary predicate P of L_1 , a formula $\psi_P(x_1 \dots x_k)$ of L_2 with exactly n free variables,
- for each n -ary function symbol F of L_1 , a formula $\psi_F(x_1 \dots x_k, y)$ of L_2 with exactly $k + 1$ free variables such that

$$S \vdash \bigwedge_{i=1}^k \chi(x_i) \rightarrow (\exists! y)(\chi(y) \& \psi_F(x_1 \dots x_k, y))$$

and similarly for ψ_c ($S \vdash (\exists! y)(\chi(y) \& \psi_c(y))$).

Using ψ_P , ψ_F and ψ_c we may define in S a predicate P^* , a function F^* (defining $F^*(x_1 \dots x_k)$ by ψ_F if $\bigwedge_i \chi(x_i)$ and putting e.g. $F^*(x_1 \dots) = x_1$ otherwise) and a constant c^* ; this extends to a translation of each L_1 -formula $\varphi(x_1 \dots x_k)$ into an L_2 -formula $\varphi^*(x_1, \dots, x_k)$ (more precisely: into a formula of L_2 enriched by the $*$ -symbols); φ^* results from φ by replacing each predicate, function, constant by its starred counterpart. The formula φ^* may be called " φ in the sense of the interpretation" or "the translation of φ ".

(3) If T is a theory in the language L_1 , S a theory in L_2 and $*$ is an interpretation of L_1 in S then $*$ is an *interpretation* of T in S if, for each axiom $\varphi(x_1, \dots, x_k)$ of T ,

$$S \vdash \bigwedge_{i=1}^k \chi(x_i) \rightarrow \varphi^*.$$

We have the following evident theorem:

1.3 Theorem. If $*$ is an interpretation of T in S then for each L_1 formula φ , $T \vdash \varphi$ implies $S \vdash \bigwedge \chi(x_i) \rightarrow \varphi^*$; in particular, for each closed L_1 -formula φ , $T \vdash \varphi$ implies $S \vdash \varphi^*$.

Proof by induction on the length of a proof. □

1.4 Definition. This generalizes to a *parametrical interpretation* of L_1 in S : it consists of a formula $\vartheta(\mathbf{z})$ such that $S \vdash (\exists \mathbf{z})\vartheta(\mathbf{z})$ (range of parameters), a formula $\chi(x, \mathbf{z})$ such that $S \vdash \vartheta(\mathbf{z}) \rightarrow (\exists x)\chi(x, \mathbf{z})$ (range of L_1 -variables given parameters) and for each P, F, c as above formulas $\psi_P(\mathbf{x}, \mathbf{z})$, $\psi_F(\mathbf{x}, \mathbf{y}, \mathbf{z})$, $\psi_c(\mathbf{y}, \mathbf{z})$ of appropriate arities such that $S \vdash \vartheta(\mathbf{z}) \& \bigwedge_i \chi(x_i, \mathbf{z}) \rightarrow (\exists! y)(\chi(y, \mathbf{z}) \& \psi_F(\mathbf{x}, \mathbf{y}, \mathbf{z}))$ and similarly for ψ_c . The definition of φ^* for a formula $\varphi(x_1 \dots x_n)$ (not containing the variables $\mathbf{z}$) is clear; $*$ is a *parametrical interpretation* of T in S if, for each axiom $\varphi(\mathbf{x})$ of T ,

$$S \vdash (\vartheta(\mathbf{z}) \& \bigwedge_i \chi(x_i, \mathbf{z})) \rightarrow \varphi^*(\mathbf{x}, \mathbf{z}).$$

1.5 Remark. (1) A generalization of 1.3 for parametrical interpretation is evident.

(2) The (parametrical) interpretation $*$ has *absolute equality* if $S \vdash x = *y \rightarrow x = y$. We restrict ourselves to interpretations with absolute equality except when stated otherwise.

(3) Note that if $*$ is parametrical and S can define an object satisfying ϑ (i.e. for some $\vartheta'(z)$, $S \vdash \vartheta'(z) \rightarrow \vartheta(z)$ and $S \vdash (\exists!z)\vartheta'(z)$ then $*$ may be replaced by a non-parametrical interpretation $**$ such that, for each closed T -formula φ , $S \vdash \varphi^* \rightarrow \varphi^{**}$ (just postulate that z is the unique tuple satisfying ϑ' throughout). This is in particular the case if S contains a fragment of arithmetic and $\vartheta(z)$ is an arithmetical formula such that S proves $L\vartheta$ (least number principle).

(4) The reader may observe that if $*$ is an interpretation of T in S then $*$ defines in each model $M \models S$ a model $M' \models T$; similarly for a parametrical interpretation.

1.6 Corollary. If T is (parametrically) interpretable in S and S is consistent then T is also consistent.

1.7 Remarks. (1) Clearly, each theory T is interpretable in itself by means of the *identical interpretation*.

(2) If T is interpretable in S and S is interpretable in U then T is interpretable in U by means of the *composed interpretation*.

1.8 Lemma. If T is interpretable in $(S + \{\varphi\})$ and also in $(S + \{\neg\varphi\})$ then T is interpretable in S .

Proof. Let $(\chi_i, \psi_{P,i} \dots)$ be the interpretation in question; ($i = 1, 2$). Put $\chi(x) \equiv (\varphi \& \chi_1(x)) \vee (\neg\varphi \& \chi_2(x))$,

$$\psi_P(\mathbf{x}) \equiv (\varphi \& \psi_{P,1}(\mathbf{x})) \vee (\neg\varphi \& \psi_{P,2}(\mathbf{x}))$$

etc. Similarly for parametrical interpretations. □

1.9 Definition. An interpretation $*$ of T in S is Γ -faithful if, for each closed $\varphi \in \Gamma$, $T \vdash \varphi$ is equivalent to $S \vdash \varphi^*$. It is *faithful* if it is Γ -faithful for Γ being the set of all L_1 -formulas.

(b) Theories Containing Arithmetic; Sequential Theories; PA and ACA_0

1.10 Convention. Saying that S contains T we shall mean that T is interpretable in S and a certain interpretation of T in S has been fixed. Thus, in particular, if T is a subtheory of S then S contains T ; but in general axioms of T are assumed to hold only on a subdomain of the universe of S . A typical example is Zermelo-Frankel set theory ZF : in ZF we can define a set N and prove that its elements satisfy axioms of PA with respect to appropriately defined operations of successor, addition and multiplication. In other words, PA is interpreted in ZF by these definitions.

1.11 Remark. It is possible that there are two substantially different interpretations of T in S , so we have to fix one.

1.12 Definition. A theory T containing Q is *sequential* if there are predicates $SEQ(z, u)$ and $\beta(x, v, z)$ coding sequences of arbitrary objects of T in the following weak sense (read $SEQ(z, u)$ “ z codes a sequence of length at least u ” and $\beta(x, v, z)$ “ x is the v -th element of z ”; let $Number(u)$ be the domain of the chosen interpretation of Q in T): T proves

- (1) $SEQ(z, u) \rightarrow Number(u) \ \& \ (\forall v < u)(Number(v) \rightarrow (\exists! x)\beta(x, v, z));$
- (2) $SEQ(z, u) \rightarrow (\forall y)(\exists z')(SEQ(z', u+1) \ \& \ (\forall v < u)(Number(v) \rightarrow (\forall x)(\beta(x, v, z) \equiv \beta(x, v, z')) \ \& \ \beta(y, u, z'))).$

(Here $u+1$ means the successor of u in the sense of the interpretation.)

1.13 Remark. (1) says that if z codes a sequence of length u then for each $v < u$, the v -th element of z is uniquely determined; and (2) is a prolongation axiom. Note that we do not define the length of a sequence. Nor do we have any extensionality etc. This notion is particularly important for the case that T has also objects that are not numbers, as the theory ACA_0 below.

Note that a theory T in the language of arithmetic stronger than $I\Sigma_0^{exp}$ is sequential; we may take $Seq(s) \ \& \ lh(s) = u$ for $SEQ(s, u)$ and $(s)_v = x$ for $\beta(x, v, s)$. In Chap. V we show that each theory in the language of arithmetic containing $I\Sigma_0$ is sequential.

In the next definition we introduce a particular system ACA_0 of second order arithmetic and later we show that it contains PA . Results of this chapter will have various corollaries concerning relations of PA and ACA_0 . The reader uninterested in ACA_0 may skip the rest of the subsection.

1.14 Definition. The language of second order arithmetic consists of

- (1) variables of two sorts: number variables $x, y, \dots$, and set variables $X, Y, \dots$.
- (2) predicates $=, \leq, \in$ (binary), function symbols $S, +, *$ (usual arities) and a constant $\bar{0}$.

First order terms coincide with terms of the language of (first order) arithmetic, i.e. are elements of the free algebra generated by number variables, the constant $\bar{0}$ and function symbols. *Second order terms* are set variables.

Atomic formulas are formulas of the following form: $t = s$, $t \in X$, $X = Y$ where t, s are first order terms and X, Y are second order variables. *Formulas* result from atomic formulas using connectives and first order and second order quantifiers $(\forall x)$, $(\forall X)$.

The second order arithmetic ACA_0 has the language of second order arithmetic and the following axioms:

- (1) Axioms of Robinson's arithmetic Q ,
 (2) Arithmetical comprehension scheme: for each formula $\varphi(x, y, X)$ *not containing any second order quantifiers*, and not containing the variable Z , the axiom

$$(\exists Z)(\forall x)(x \in Z \equiv \varphi(x, y, X))$$

- (3) The following (single) *induction axiom*:

$$\bar{0} \in X \ \& \ (\forall x)(x \in X \rightarrow S(x) \in X) \rightarrow (\forall x)(x \in X).$$

- (4) $X = Y \equiv (\forall x)(x \in X \equiv x \in Y)$ (extensionality).

1.15 Remarks. (1) How does this fit into our notion of an axiomatic theory? It is indeed possible to generalize the notion of an axiomatic theory as presented in Sect. 0 to allow various sorts of variables (see e.g. [Kreisel-Krivine]); but, on the other hand, we may consider any many-sorted theory as one-sorted with an unary predicate for each sort. This means for ACA_0 : Instead of variables of two sorts we have two predicates $Number(x)$, $Set(x)$ and axioms that there are numbers, there are sets and $Number(x) \equiv \neg Set(x)$; we have predicates and function symbols of the language L_0 of first order arithmetic and in addition $\in$. We formulate axioms saying that numbers are closed under S , $+$, $*$ and satisfy Q (e.g. $Number(x) \ \& \ x \neq 0 \rightarrow (\exists y)(Number(y) \ \& \ x = S(y))$ etc.). Similarly we rewrite the comprehension.

$$Number(y) \ \& \ Set(w) \rightarrow (\exists z)(Set(z) \ \& \ (\forall x)(Number(x) \rightarrow (x \in z \equiv \varphi'(x, y, w))))$$

(where φ' results from a formula φ above by the obvious changes) and the induction axiom. Finally we add extensionality:

$$Set(x) \ \& \ Set(y) \rightarrow (\forall u)(Number(u) \rightarrow (u \in x \equiv u \in y)) \equiv x = y.$$

Thus we understand the two-sorted formulation as a shorthand for the one-sorted system just described. (The interested reader could show that this is equivalent to a consequent many-sorted approach; models of both formulations are in an obvious correspondence.)

(2) How does ACA_0 fit into various systems of second order arithmetic? And why is ACA_0 introduced and studied here? A much stronger theory results from ACA_0 if we postulate comprehension for *all* φ ; this is *the full* second order arithmetic. It has important particular subsystems studied in Simpson's book [Simpson]. We shall not go into this; but we shall show below that ACA_0 extends PA conservatively and ACA_0 is finitely axiomatizable. We shall see later in this chapter that PA is *not* finitely axiomatizable; but it is the union of the hierarchy of theories $I\Sigma_k$, $k \in N$, each being finitely axiomatizable (see Chap. I, Sect. 2) and has a finitely axiomatizable

conservative extension (in a richer, second-order language), namely ACA_0 . Note in passing that the relation of ACA_0 to PA is very similar to the relation of Gödel-Bernays set theory GB to Zermelo-Fraenkel set theory ZF (and, of course, both ZF and GB contain both PA and ACA_0).

1.16 Theorem. ACA_0 is a conservative extension of PA ; more precisely, the obvious interpretation of the language of PA in ACA_0 is a faithful interpretation of PA in ACA_0 .

Remark. If we treat ACA_0 consequently as a two-sorted theory then formulas of PA are particular formulas of ACA_0 and it makes sense to say that ACA_0 is a conservative extension of PA . If we understand ACA as a one sorted theory, we have to interpret PA -formulas as ACA_0 -formulas by restricting all quantifiers to $Number(x)$. But if there is no danger of misunderstanding we shall identify both approaches.

Proof. First we show that ACA_0 extends PA , i.e. that each induction axiom I_φ is provable in ACA_0 . But this is easy: by comprehension, let $Z = \{x \mid \varphi(x, y)\}$. Assume $\varphi(0, y)$ and $(\forall x)(\varphi(x, y) \rightarrow \varphi(S(x), y))$. Then $0 \in Z$ and $(\forall x)(x \in Z \rightarrow S(x) \in Z)$. We get $((\forall x)(x \in Z))$, i.e. $(\forall x)\varphi(x, y)$. Now we prove conservativity.

A model-theoretic proof is easy: take any countable model M of PA ; without loss of generality you may assume that no subset of M is an element of M . Interpret elements of M as numbers and take all parametrically definable subsets of M for sets: Let S be the set of all such subsets. The new model M' has the domain $M \cup S$, operations and ordering are as in M (trivially extended to S), and $\in$ is interpreted as the restriction of actual membership to $M \times S$. Checking that $M' \models ACA_0$ is not difficult and is left to the reader. (One may consult 0.9 and/or Chap. IV, Sect. 1 (a) if necessary). Thus: if a PA -sentence φ is consistent with PA it is consistent with ACA_0 ; thus ACA_0 is conservative over PA . $\square$

Remark. In Chap. IV, Sect. 4 we shall formalize this proof in IS_1 and show that the assertion “ ACA_0 extends PA conservatively” is provable in IS_1 . By the results of Chap. IV, this will imply that there is a *primitive recursive* function associating to each ACA_0 -proof of a PA -formula its PA -proof. (Better results are known.)

1.17 Theorem. ACA_0 is sequential.

Proof. We have to define coding of sequences of numbers and sets. We shall define a coding of sequences of sets, and give indications how to modify the definition to code sequences of both numbers and sets.

Define $SEQ(Z, u)$ iff Z consists of ordered pairs; and put $\beta(X, v, Z)$ iff $(\forall x)(x \in X \equiv (v, x) \in Z)$. Note that if $SEQ(Z, u)$ then $SEQ(Z, u')$ for any

u' . Thus we write just $SEQ(Z)$. If $SEQ(Z)$ and Y is a class then define $G(Z, Y, u)$ as follows:

$$(x, v) \in G(Z, Y, u) \equiv (v \neq u \& (x, v) \in Z \vee v = u \& y \in Y).$$

Existence follows by comprehension; and verification of necessary provabilities is easy.

If we want to code sequences of both sets and numbers we may consider sets Z of triples $((x, \varepsilon), v)$ such that for each v either for all (x, ε) such that $((x, \varepsilon), v) \in Z$ we have $\varepsilon = 1$ (and then $\beta(X, v, Z) \equiv X = \{x \mid ((x, 1), v) \in Z\}$ or there is exactly one x such that $(x, 0), v) \in Z$ (and then $\beta(x, v, Z) \equiv ((x, 0), v) \in Z$). $\square$

1.18 Theorem. ACA_0 is finitely axiomatizable.

Proof. We are interested in a quick proof, not in a polished finite axiom system. Thus we shall start with $I\Sigma_1$, which we know to be finitely axiomatizable. We add the induction axiom (A1), which is a single axiom. Observe that ACA_0 proves $(\forall X)(X \text{ is piecewise coded})$ (consider the class of all x such that there is a piece of X of length x); thus take the axiom (A2) saying that each X is p.c. Recall the relativized satisfaction $Sat_{0,X}$ (see I.2.55); take the axiom (A3) $(\forall f \in \Sigma_0^*(X))(\exists Z)(\forall e)(e \in Z \equiv Sat_{0,X}(f, e))$ (we may desire that free variables of f are the first u free variables and e is an u -tuple of numbers). In fact, it is enough to quantify f over all *open* $\Sigma_0(X)$ -formulas. Note that our theory proves Tarski's truth conditions for $Sat_{0,X}$ so that we have "snowing"-snowing for $\Sigma_0^*(X)$ -formulas. (A4) and (A5) will describe universal and existential projection:

$$(A4) (\forall X)(\exists Z)(\forall s)(s \in Z \equiv (\forall x)(\langle x \rangle \frown s \in X))$$

$$(A5) (\forall X)(\exists Z)(\forall s)(s \in Z \equiv (\exists x)(\langle x \rangle \frown s \in X))$$

$$(A6) \text{ says } (\forall X)(\exists Z)(\forall x)(x \in Z \equiv \langle x \rangle \in X).$$

Clearly, $I\Sigma_1 + (A2) - (A6)$ proves each instance of comprehension for a formula $\varphi(x, X)$ without second order quantifier: Apply (A3) to φ_0 , where φ_0 is the open part of φ (φ assumed prenex) and then apply (A4) and (A5) according to the quantifier prefix of φ . Thus in particular we have comprehension for all formulas not containing any set variables at all.

We add one more axiom that reduces the general case to the subcase $\varphi(x, X)$ just described.

We take the axiom stating the property of the function $G(Z, Y, u)$:

$$\begin{aligned} (\forall u)(\forall Y)(\forall Z)(SEQ(Z) \rightarrow (\exists W)(\forall x, v)((x, v) \in W \\ \equiv (v \neq u \& (x, v) \in Z \vee v = u \& x \in Y)). \end{aligned}$$

This makes possible to replace k sets by just one; e.g. X_0, X_1, X_2 are replaced by $G(G(G(0, X_0, 0), X_1, 1), X_2, 2)$. $\square$

(c) Numerations and Binumerations

We have already mentioned the notion of a binumeration of a set of natural numbers in a theory $T \supseteq Q$ (see I.1.65). In this short subsection we define a more general notion of a numeration and prove some easy facts on them. Deeper theorems on numerations will be proved (and used) in Sect. 3. At the end of this subsection we shall show that our present knowledge is sufficient for a proof of a weak form of Gödel's first incompleteness theorem. A strong form will be proved by means of self-reference in Sect. 2.

1.19 Definition. Let $R \subseteq N^n$ be a relation. A formula $\varphi(x_1, \dots, x_n)$ *numerates* R in a theory T containing Q if, for each $k_1, \dots, k_n \in N$, we have the following:

$$(*) \quad R(k_1, \dots, k_n) \Leftrightarrow T \vdash \varphi(\bar{k}_1, \dots, \bar{k}_n).$$

1.20 Remark. Note that φ binumerates R in T if φ numerates R in T and $\neg\varphi$ numerates the complement of R in T , i.e. besides $(*)$ we have

$$\text{not } R(k_1, \dots, k_n) \Leftrightarrow T \vdash \neg\varphi(k_1, \dots, k_n).$$

1.21 Definition. A theory T containing Q is Γ -*sound* (where Γ is a class of formulas of L_0) if each L_0 -formula provable in T is true in the standard model N . T is *sound* if it is Γ -sound for Γ being the class of all L_0 -formulas. (Thus PA is sound and so are its subsystems.)

1.22 Remark. We shall be particularly interested in Σ_1 -sound theories. They are also called 1-consistent; T is 1-consistent if $(T + Tr(\Pi_1))$ is consistent, where $Tr(\Pi_1)$ is the set of all Π_1 -formulas true in N . (Show that the two definitions are equivalent.) Note that Σ_1 -soundness implies consistency.

1.23 Theorem. (1) If A is defined by a Σ_1 -formula $\varphi(x_1, \dots, x_n)$ in N then φ numerates A in Q and also in each Σ_1 -sound theory T containing Q .

(2) If A is Δ_1 (in N) then there is a Σ_1 -formula φ which binumerates A in Q and also in each consistent T containing Q .

Proof. (1) follows immediately from Σ_1 -completeness and Σ_1 -soundness. (See I.1.8). To prove (2) first assume $A \subseteq N$; take a Σ_1 -definition $(\exists y)\sigma(x, y)$ and a Π_1 -definition $(\forall y)\pi(x, y)$ of A ($\sigma, \pi \in \Sigma_0$). Consider the following formula $\varphi(x)$:

$$\varphi(x) \equiv (\exists y)(\sigma(x, y) \& (\forall z \leq y)\pi(x, z)).$$

(This is our first example of *witness comparison* that will play an extremely important role in the sequel: The formula says that there is a witness y for

$(\exists u)\sigma(x, u)$ such that beneath y there is no witness for $(\exists u)\neg\pi(x, u)$. If $k \in A$ then for some m , $N \models \sigma(\bar{k}, \bar{m})$, trivially, for each $m' \leq m$, $N \models \pi(\bar{k}, \bar{m})$ since $N \models (\forall y)\pi(k, y)$. Thus Q proves $\neg\pi(\bar{k}, \bar{m}) \& (\forall y \leq \bar{m})\neg\sigma(k, y)$. But then Q proves $\neg\varphi(\bar{k})$: Work in Q and assume $\varphi(\bar{k})$; let y_0 be such that $\sigma(\bar{k}, y_0)$ and $(\forall z \leq y_0)\pi(k, z)$. We have either $y_0 \leq \bar{m}$ or $y_0 \geq \bar{m}$ (cf. I.1.6). But $y_0 \leq \bar{m}$ implies $\neg\sigma(\bar{k}, y_0)$, a contradiction, and $y_0 \geq \bar{m}$ implies $\pi(\bar{k}, \bar{m})$, also a contradiction. This completes the proof. $\square$

Now assume that R is an n -ary relation. Let ML_n be the maximo-lexicographical ordering of N^n , i.e. $(k_1, \dots, k_n)ML_n(q_1, \dots, q_n)$ iff $\max(k_1, \dots, k_n) < \max(q_1, \dots, q_n)$ or [the maxima are equal and for the first i such that $k_i \neq q_i$ we have $k_i < q_i$]. Define ML_n by a Σ_0 formula $(x_1, \dots, x_n) \leq_n (y_1, \dots, y_n)$ by copying the definition (caution: we use a *disjunction* over $i = 1, \dots, n$). Analogously as for $<$, Q proves $(x_1, \dots, x_n) \leq_n (\bar{q}_1, \dots, \bar{q}_n) \equiv \bigvee \{ \bigwedge_i x_i = \bar{k}_i \mid (\bar{k}_1, \dots, \bar{k}_n)ML_n(\bar{q}_1, \dots, \bar{q}_n) \}$. (Cf. 5.1.6.). Having this construct a $\varphi(x_1, \dots, x_n)$ from $(\exists u)\sigma(x_1 \dots x_n, u)$ and $(\forall u)\pi(x_1, \dots, x_n, u)$ as above but using $\leq_n$ instead of $<$.

1.24 Remark. (1) The preceding theorem implies that each Δ_1 relation A also has a Π_1 binumeration in $T \supseteq Q$ (since the complement of A is also Δ_1)

(2) Recall the simple remark in I.1.65 which states that if A is defined by a formula φ which is Δ_1 in T then A is binumerated by φ in T .

(3) In Sect. 3 we remove the assumption of Σ_1 soundness from 1.23 (1).

(4) Functions are particular relations; thus if F is a Δ_1 mapping of N into N then there is a Σ_1 formula φ such that

$$\begin{aligned} k = F(m) &\Leftrightarrow Q \vdash \varphi(\bar{k}, \bar{m}) \\ k \neq F(m) &\Leftrightarrow Q \vdash \neg\varphi(\bar{k}, \bar{m}) \end{aligned}$$

But for functions we can say even more (generalization for functions of several variables being left to the reader):

1.25 Theorem. Let $F : N \rightarrow N$ be Σ_1 (possibly partial). Then there is a Σ_1 formula $\varphi(x, y)$ such that, for each $m \in \text{dom}(F)$,

$$Q \vdash \varphi(\bar{m}, y) \equiv y = \overline{F(\bar{m})}.$$

Proof. Let F be defined by a Σ_1 -formula $(\exists z)\sigma(x, y, z)$, σ being Σ_0 (i.e. $F(m) = k$ iff $N \models (\exists z)\sigma(\bar{m}, \bar{k}, z)$). Let $\varphi(x, y)$ be the following formula $(\exists z)\varphi_0(x, y, z)$:

$$\begin{aligned} (\exists z)[\sigma(x, y, z) \& (\forall u, v \leq y)(u \neq y \rightarrow \neg\sigma(x, u, v) \\ \& (\forall u, v \leq z)(u \neq y \rightarrow \neg\sigma(x, u, v))]. \end{aligned}$$

(Note that in general we cannot speak on $\max(y, z)$ in Q .) First let $F(m) = k$, $N \models \sigma(\bar{m}, \bar{k}, \bar{q})$; then it is easy to show $Q \vdash \varphi_0(\bar{m}, \bar{k}, \bar{q})$. Thus $Q \vdash y = \bar{k}_1 \rightarrow \varphi(\bar{m}, y)$. Second, work in Q and assume $u \neq \bar{k}$ and let $\varphi_0(\bar{m}, u, v)$. For k and q we do have a maximum; denote it h .

Case 1. $\bar{h} \leq u$ or $\bar{h} \leq v$. Then start from $\varphi_0(\bar{m}, u, z)$ and show $\neg\sigma(\bar{m}, \bar{k}, \bar{q})$ – contradiction.

Case 2. $u, v \leq \bar{h}$. Then start from $\varphi_0(\bar{m}, \bar{k}, \bar{q})$ and show $\neg\sigma(\bar{m}, u, v)$, again a contradiction. This shows (in Q) $y \neq \bar{k} \rightarrow \neg\varphi(\bar{m}, y)$.

Remark. The reader may formulate and prove a generalization to functions with several arguments as an exercise.

1.26 Remark. In Sect. 2 we shall study Gödel's celebrated incompleteness theorem; our main method will be Gödel's self-reference technique. In the following remark we shall show how to give a non-constructive proof of a weak form of Gödel's first incompleteness theorem using means available so far. We claim the following:

If T is a Σ_1 -sound Σ_1 theory containing Q then T is incomplete; there is a Σ_1 sentence ν such that both ν and $\neg\nu$ are unprovable in T .

Proof. Recall that there is a Σ_1 set K_0 of natural numbers which is not Δ_1 (the reader may find a proof of this in the next remark).

Now let $\alpha(x)$ be a Σ_1 formula numerating K_0 in T , i.e. $K_0 = \{n \mid T \vdash \alpha(\bar{n})\}$. Let $K_1 = \{n \mid T \vdash \neg\alpha(\bar{n})\}$. Since K_0 is not Δ_1 and T is consistent we have $K_0 \cup K_1 \neq N$; for $n \in N - K_0 - K_1$ both $\alpha(\bar{n})$ and $\neg\alpha(\bar{n})$ are unprovable in T . $\square$

1.27 Remark. For the reader's convenience we prove here that there is a Σ_1 set of natural numbers which is not Δ_1 . For example, let K_0 be the set of all φ such that φ is a Σ_1 -formula with just one free variable and is satisfied on N by φ (itself). Then K_0 is Σ_1 (it is defined by

$$x \in \Sigma_1^* \ \& \ x \text{ has one free variable}^* \ \& \ Sat_{\Sigma,1}(x, [x]).$$

If $N - K$ were Σ_1 then we could produce a Σ_1 -formula $\varphi(x)$ defining the same set of natural numbers as the formula

$$x \in \Sigma_1^* \ \& \ x \text{ has one free variable}^* \ \& \ \neg Sat_{\Sigma,1}(x, [x]).$$

By “snowing”-snowing.

$$N \models \bar{\varphi} \in \Sigma_1^* \ \& \ \bar{\varphi} \text{ has one free variable}^* ;$$

now it is easy to see that we get the liar's paradox:

$$N \models Sat_{\Sigma,1}(\bar{\varphi}, [\bar{\varphi}]) \equiv \neg Sat_{\Sigma,1}(\bar{\varphi}, [\bar{\varphi}]),$$

which is a contradiction. We have shown that K_0 is not Σ_1 .

2. Self-Reference and Gödel's Theorems, Reflexive Theories

We now come to the substance of Gödel's method of proof of incompleteness of arithmetic, namely to the use of arithmetization for *self-reference*, which is often roughly described as the existence of arithmetical sentences "speaking of themselves". Since in our representation, sentences are particular numbers, we are not surprised if we encounter sentences speaking of (other) sentences. Take, for example, the case of "snowing"-snowing: we know that $N \models \varphi$ iff $N \models \text{Sat}_{\Sigma, n}(\bar{\varphi}, \emptyset)$ (φ being Σ_n). Given any formula $\psi(x)$ with exactly one free variable, a sentence φ is *self-referential* with respect to ψ in N (or: is a *fixed-point* of ψ in N) if $N \models \varphi \equiv \psi(\bar{\varphi})$, thus $N \models \varphi$ is equivalent to $N \models \psi(\bar{\varphi})$. We can say rather suggestively that φ says "I have the property ψ ". As usual, we are more interested in provability in a theory than in truth in N ; we say that φ is a *fixed-point of $\psi(x)$ in T* if $T \vdash \varphi \equiv \psi(\bar{\varphi})$.

In subsection (a) we prove various theorems on the existence of fixed points; in (b) we prove Gödel's incompleteness theorems and related results, and in (c) we derive consequences for theories similar to full Peano arithmetic PA ; among other things we show that PA is not finitely axiomatizable. We shall pay much attention to finitely axiomatizable theories in Sect. 3.

(a) Existence of Fixed Points

2.1 Fixed-point Theorem (Or Diagonal Lemma).

- (1) ((Non-parametric version). Let T be a theory containing Q and let $\psi(x)$ be a formula with exactly one free variable. Then there is a sentence φ such that $T \vdash \varphi \equiv \psi(\bar{\varphi})$.
- (2) (Parametric version). Let $T \supseteq Q$ and let $\psi(x, y)$ be a formula with free variables as indicated. Then there is a formula $\varphi(y)$ such that $T \vdash \varphi(y) \equiv \psi(\bar{\varphi}(y), y)$.

Remark. We prove (1) and indicate how to generalize to (2). The rest of the subsection contains some corollaries and related results as well as definitions of some properties that fixed points may have or not have.

Proof of (1). Let $\psi(x)$ be given. Let F be the Δ_1 function associating with each formula $\delta(x)$ (in the language of T having exactly one free variable x) the closed formula $\delta(\bar{\delta})$, i.e. the result of substitution of $\bar{\delta}$ into δ for x . (Let F be 0 for other arguments.) By 1.25 there is a formula $\alpha(x, v)$ such that, for each δ , $T \vdash \alpha(\bar{\delta}, v) \equiv v = F(\bar{\delta})$. Let $\chi(x)$ be the formula $(\exists v)(\alpha(x, v) \& \psi(v))$ and let φ be $F(\chi)$, i.e. $\chi(\bar{\chi})$. Q proves the following equivalences:

$$\begin{aligned} \varphi \equiv \chi(\bar{\chi}) &\equiv (\exists v)(\alpha(\bar{\chi}, v) \& \psi(v)) \equiv (\exists v)(v = F(\bar{\chi}) \& \psi(v)) \\ &\equiv (\exists v)(v = \bar{\varphi} \& \psi(v)) \equiv \psi(\bar{\varphi}). \end{aligned}$$

To prove (2) modify F . F associates with each formula $\delta(x, y)$ the formula $\delta(\bar{\delta}, y)$. Let α be as above, for the new F . If $\psi(x, y)$ is given, let $\chi(x, y)$ be the formula $(\exists v)(\alpha(x, v) \& \psi(v, y))$, let φ be $F(\chi)$. We have

$$Q \vdash \varphi(y) \equiv \psi(\bar{\varphi}, y)$$

as above. □

2.2. Corollary. (1) Let T contain Q and let $\psi(x, z)$ be a T -formula. Then there is a T -formula $\varphi(z)$ such that for each $k \in N$, $T \vdash \varphi(\bar{k}) \equiv \psi(\overline{\varphi(\bar{k})}, \bar{k})$.

(2) More generally, if $\psi((x, z, y))$ is a T -formula then there is a T -formula $\varphi(z, y)$, such that, for each k ,

$$T \vdash \varphi(\bar{k}, y) \equiv \psi(\overline{\varphi(\bar{k}, y)}, \bar{k}, y).$$

Proof. We prove (1). Let $F(\delta, k)$ be the Δ_1 function associating with each k the formula $\delta(\bar{k})$; let $\beta(u, v, w)$ be such that $Q \vdash \beta(\bar{\delta}, \bar{k}, \bar{w}) \equiv w = \overline{F(\delta, k)}$. Let $\psi'(x, z)$ be $(\exists w)(\beta(x, z, w) \& \psi(w, z))$; let $\varphi(z)$ be such that

$$T \vdash \varphi(z) \equiv \psi'(\bar{\varphi}, z).$$

Then $T \vdash \varphi(\bar{k}) \equiv \psi'(\bar{\varphi}, \bar{k}) \equiv (\exists w)(\beta(\bar{\varphi}, \bar{k}, w) \& \psi(w, \bar{k}) \equiv \psi(\overline{\varphi(\bar{k})}, \bar{k}))$. □

The proof of the parametric version is left to the reader as an easy exercise.

2.3 Corollary. Let T be a consistent theory containing Q . Then there is no formula $Tr(x)$ such that for each φ , $T \vdash \varphi \equiv Tr(\bar{\varphi})$.

Proof. If we had $Tr(x)$ we could reproduce the liar's paradox: let φ be such that $T \vdash \varphi \equiv \neg Tr(\bar{\varphi})$ (by 2.1). Then $T \vdash \varphi \equiv \neg \varphi$, i.e. T is inconsistent. □

2.4 Remark. This means that T cannot define its own truth. In particular, if $T \supseteq I\Sigma_1$ is a consistent theory in the language of arithmetic we have, for each k , partial satisfactions $Sat_{\Sigma, k}$ and $Sat_{\Pi, k}$ for the universe but there is no formula $Sat(z, e)$ such that T proves Tarski's properties of satisfaction for all formulas.

2.5 Discussion. Let $T \vdash (\forall x)\psi(x)$; then φ is a fixed point of ψ in T iff $T \vdash \varphi$. (We can say that ψ has only provable fixed points.). If $T \vdash (\forall x)\neg\psi(x)$ then ψ has only refutable fixed points. In both cases, all fixed points of ψ are T -provably equivalent, i.e. if φ_1 and φ_2 are fixed points of φ then $T \vdash \varphi_1 \equiv \varphi_2$. We say that ψ has a unique fixed point up to an equivalence. These are trivial examples; we shall construct non-trivial examples in the next subsection.

2.6 Theorem (Existence of Self-referential Pairs). Let T contain Q and let $\psi_1(x, y), \psi_2(x, y)$ be two T -formulas. Then there are T -sentences φ_1, φ_2 such that $T \vdash \varphi_1 \equiv \psi_1(\bar{\varphi}_1, \bar{\varphi}_2)$ and $T \vdash \varphi_2 \equiv \psi_2(\bar{\varphi}_1, \bar{\varphi}_2)$.

Proof. Let $F_i(\delta_1(x, y), \delta_2(x, y))$ be $\delta_i(\bar{\delta}_1, \bar{\delta}_2)$ ($i = 1, 2$); let $\alpha_i(u, v, w)$ be such that $Q \vdash \alpha_i(\bar{\delta}_1, \bar{\delta}_2, w) \equiv w = F_i(\delta_1, \delta_2)$. Let $\chi(x, y)$ be $(\exists w_1, w_2)(\alpha_1(x, y, w_1) \& \alpha_2(x, y, w_2) \& \psi_i(w_1, w_2))$; let φ_i be $\chi_i(\bar{\chi}_1, \bar{\chi}_2)$. The proof of $T \vdash \varphi_i \equiv \psi_i(\bar{\varphi}_1, \bar{\varphi}_2)$ is now routine. $\square$

(b) Gödel's First Incompleteness Theorem and Related Topics

Fixed points theorems are particularly fruitful when applied to formulas related to the set of all T -proofs or the set of all T -provable formulas (where T is a theory containing Q). If T is rich enough to make arithmetization possible ($T \supseteq I\Sigma_1$ suffices) then it is natural to work with the formalized proof predicate $Proof^*$ (see Chap. I, Sect. 4); but for Gödel's first incompleteness theorem it suffices that T contain Q and that the set of all proofs is Δ_1 and hence has a Σ_1 binumeration. (Gödel's second incompleteness theorem is more delicate and to prove it we have to guarantee some provability conditions, see below.) The theorems are proved in a constructive way, i.e. we give *examples* of sentences that are independent (neither provable nor refutable). Since these examples are both famous and useful we give their names (Gödel's formula, Rosser's formula).

2.7 Definition. Let T be a Σ_1 theory containing Q and let $\pi(x)$ be a Σ_1 formula defining the set of all T -provable sentences (i.e. $T \vdash \varphi$ iff $N \models \pi(\bar{\varphi})$). A *Gödel sentence* based on π is a fixed point of $\neg\pi$, i.e. a sentence ν such that

$$T \vdash \nu \equiv \neg\pi(\bar{\nu}).$$

(We may describe ν by saying that it asserts its own unprovability.)

2.8 Theorem. Let T, π be as above and let ν be a Gödel sentence given by π .
(1) If T is consistent then T does not prove ν . (2) If T is Σ_1 -sound then T does not prove $\neg\nu$.

Proof. (1) Assume $T \vdash \nu$, then $N \models \pi(\bar{\nu})$, thus $T \vdash \pi(\bar{\nu})$ by Σ_1 -completeness; thus $T \vdash \neg\nu$ and T is inconsistent.

(2) Assume $T \vdash \neg\nu$, i.e. $T \vdash \pi(\bar{\nu})$. If T were Σ_1 -sound then π would *numerate* the set of all provable sentences, thus we would have $T \vdash \nu$ and T would be inconsistent. Thus T cannot be Σ_1 -sound. $\square$

We present a technique that makes it possible to get rid of Σ_1 -soundness.

2.9 Definition. Let $(\exists u)\alpha(x, u)$ and $(\exists u)\beta(x, u)$ be two Σ_1 -sentences; thus α, β are Σ_0 . A *witness comparison sentence* given by these sentences is the sentence

$$(*) \quad (\exists u)[\beta(x, u) \& (\forall v < u)\neg\alpha(x, u)]$$

(Observe that this is again a Σ_1 sentence; if we call each u such that $\alpha(x, u)$ a *witness* for $(\exists u)\alpha(x, u)$ and similarly for $(\exists u)\beta(x, u)$ then $(*)$ says "there is a witness u for $(\exists u)\beta(x, u)$ such that no witness for $(\exists u)\alpha(x, u)$ is $< u$ ". The formula $(*)$ will be often denoted $(\exists u)\beta \preccurlyeq (\exists u)\alpha$. In particular, let $(\exists u)\alpha(x, u)$ be a Σ_1 -formula defining the set of all T -provable sentences and let $(\exists u)\beta(x, u)$ be a Σ_1 -formula defining the set of all T -refutable sentences, i.e. $T \vdash \varphi \Leftrightarrow N \models (\exists u)\alpha(\bar{\varphi}, u)$ and $T \vdash \neg\varphi \Leftrightarrow N \models (\exists u)\beta(\bar{\varphi}, u)$. A *Rosser sentence* given by α and β is a formula ρ such that

$$T \vdash \rho \equiv (\exists u)\beta(\bar{\rho}, u) \preccurlyeq (\exists u)\alpha(\bar{\rho}, u),$$

i.e.

$$T \vdash \rho \equiv (\exists u)(\beta(\bar{\rho}, u) \& (\forall v < u)\neg\alpha(\bar{\rho}, v)).$$

Calling for a moment an element u such that $\alpha(x, u)$ a *demonstration* of x and u such that $\beta(x, u)$ a *refutation* of x , a Rosser sentence says: "there is a refutation of me beneath of which there is no demonstration of me".

2.10 Gödel-Rosser Incompleteness Theorem. Let T be a Σ_1 theory containing Q and let ρ be a Rosser sentence for T . If T is consistent then T proves neither ρ nor $\neg\rho$.

Proof. Assume $T \vdash \rho$, then for some d , $N \models \alpha((\bar{\rho}, \bar{d})$ and hence $T \vdash \alpha(\bar{\rho}, \bar{d})$. Now work in Q and let y be a witness for $(\exists y)\beta(\bar{\rho}, y)$ such that beneath y there is no witness for $(\exists y)\alpha(\bar{\rho}, y)$. Then $y \leq \bar{\rho} \vee \bar{d} \leq y$ (cf. I.1.6) and $\bar{d} \leq y$ is impossible since $\alpha(\bar{\rho}, \bar{d})$. We have proved $T \vdash (\exists y < \bar{d})\beta(\bar{\rho}, y)$, i.e. $T \vdash \bigvee_{e < \bar{d}} \beta(\bar{\rho}, \bar{e})$. But since T is consistent T does not prove $\neg\rho$ and thus for each e , $N \models \neg\beta(\bar{\rho}, \bar{e})$, thus $T \vdash \neg\beta(\bar{\rho}, \bar{e})$ by Σ_1 -completeness. Thus $T \vdash \bigwedge_{e < \bar{d}} \neg\beta(\bar{\rho}, \bar{e})$ and T is inconsistent.

Second, assume $T \vdash \neg\rho$, thus

$$T \vdash (\forall y)(\beta(\bar{\rho}, y) \rightarrow ((\exists z < y)\alpha(\bar{\rho}, z))).$$

(Beneath each refutation of me there is a demonstration of me.) There is a d such that $T \vdash \beta(\bar{\rho})$; thus $T \vdash (\exists z < \bar{d})\alpha(\bar{\rho}, z)$. Similarly as above we get $T \vdash \bigvee_{e < \bar{d}} \alpha(\bar{\rho}, \bar{e})$ and $T \vdash \bigwedge_{e < \bar{d}} \neg\alpha(\bar{\rho}, \bar{e})$, thus T is inconsistent. $\square$

2.11 Theorem (Essential Undecidability of Arithmetic). Each consistent theory T containing Q is undecidable; i.e. the set of its theorems is not Δ_1 .

Proof. This is because if T were decidable we could extend it to a Δ_1 theory T' which is complete; but this is impossible by 2.10. (Indeed, let $\varphi_0, \varphi_1, \dots$ be a Δ_1 enumeration of sentences of T ; define $\alpha_n = \varphi_n$ if $\text{Con}(T \cup \{\alpha_0, \dots, \alpha_{n-1}\})$ and $\alpha_n = \neg\varphi_n$ otherwise; axioms of T' are all α_n . This usual completion procedure is easily seen to be Δ_1 if T is decidable.) $\square$

2.12 Remark. There are many ways in which Gödel's incompleteness theorems can be generalized, strengthened or modified. We shall present some of them in this book. To close this subsection we prove a theorem showing the existence of a formula which is "as independent as possible" (called a flexible formula).

2.13 Definition. Let T contain Q . A formula $\varphi(x)$ with just one free variable is *flexible* over T if each elementary conjunction of formulas $\varphi(\bar{0}), \varphi(\bar{1}), \varphi(\bar{2}), \dots$ is consistent with T . In more details, let $(0)\alpha$ be $\neg\alpha$ and $(1)\alpha$ be α ; then for each string $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$, the theory

$$T, (\varepsilon_0)\varphi(\bar{0}), (\varepsilon_1)\varphi(\bar{1}), \dots, (\varepsilon_n)\varphi(\bar{n})$$

is consistent.

2.14 Lemma. Let T be a Σ_1 theory containing Q . Then there is a Σ_1 formula $\varphi(x)$ such that for each k , $(T + (\forall x)(\varphi(x) \equiv x = \bar{k}))$ is consistent. (The additional axiom says that k is the unique element satisfying φ .)

Proof. Define a function F as follows: $F(\psi) = k$ iff ψ is a T -formula with just one free variable and there is a T -proof d of $\neg(\forall x)(\psi(x) \equiv x = \bar{k})$ such that for each $d' < d$ and each $k' \leq d'$, d' is not a T -proof of $\neg(\forall x)(\psi(x) \equiv x = \bar{k}')$ (thus d is minimal possible). Clearly, F is a partial Σ_1 function. By 1.25, let $\alpha(x, y)$ be such that if $F(\psi) = k$ then $T \vdash \alpha(\bar{\psi}, y) \equiv y = \bar{k}$. By the diagonal lemma, let $\varphi(y)$ be such that $T \vdash \varphi(y) \equiv \alpha(\bar{\varphi}, y)$. We claim that $F(\varphi)$ is undefined, which means that for each k , $(T + (\forall x)(\varphi(x) \equiv x = \bar{k}))$ is consistent. Assume not and let d be the minimal possible proof as above, d proving $\neg(\forall x)(\varphi(x) \equiv x = \bar{k})$. Then $T \vdash \alpha(\bar{\varphi}, y) \equiv y = \bar{k}$, thus $T \vdash (\forall x)(\varphi(x) \equiv x = \bar{k})$ and T is inconsistent. $\square$

2.15 Theorem (on flexible formulas). Let T be Δ_1 theory containing Q . Then there is a flexible Σ_1 formula.

Proof. Let $\beta(s, i)$ be a Σ_1 formula defining the i -th element of a sequence s ; let $\varphi(x)$ be the formula from 2.14 and let $\psi(x)$ be

$$(\exists s)(\varphi(s) \& \beta(s, x) = 1).$$

Now let ε be a string $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n)$ and take $T + (\forall x)(\varphi(x) \equiv x = \bar{\varepsilon})$. This theory is consistent and proves $\psi(\bar{k}) \equiv \bar{\varepsilon}_k = 1$, i.e. proves $(\varepsilon_k)\psi(\bar{k})$ ($k = 0, \dots, n$). This completes the proof. $\square$

(c) Gödel's Second Incompleteness Theorem

Our formulation of Gödel's first incompleteness theorem was rather general; it concerned any Σ_1 definition π of all T -provable formulas. In formulating the second Gödel's incompleteness theorem we shall first present a rather general formulation (2.21) but then shall immediately present a particular case (2.22) concerning theories in which syntax has been developed (i.e. theories containing $I\Sigma_1$). We derive the following consequences: Löb's theorem (2.25), the fact that PA is not finitely axiomatizable (2.24) and a stronger fact saying that, for each $n \geq 1$, $I\Sigma_n$ is not axiomatizable using Σ_{n+2} formulas (2.27).

2.16 Definition. Let T be a theory containing Q and let π be a Σ_1 definition of the set of all T -provable formulas. The *provability conditions* for π are the following conditions:

- (1) $T \vdash \varphi$ implies $T \vdash \pi(\overline{\varphi})$,
- (2) $T \vdash \pi(\overline{\varphi}) \rightarrow \pi(\overline{\pi(\overline{\varphi})})$,
- (3) $T \vdash \pi(\overline{\varphi}) \& \pi(\overline{\varphi \rightarrow \psi}) \rightarrow \pi(\overline{\psi})$

2.17 Lemma. Let T contain $I\Sigma_1$, let τ be a Σ_1 definition of T and let the predicates $Proof_\tau^\bullet(s, x)$ (s is a τ -proof of x) and $Pr_\tau^\bullet(x)$ (x is τ -provable) be as in I.4.3. Then $Pr_\tau^\bullet$ satisfies the provability conditions 2.16.

Proof. Observe that if τ is a Σ_1 definition of T then $Pr_\tau^\bullet$ is a definition of the set of all T -provable formulas and is Σ_1 in T . Thus if $T \vdash \varphi$ then the formula $Pr_\tau^\bullet(\overline{\varphi})$ is true and therefore T -provable, by Σ_1 -completeness (see I.1.8). This proves (1). To prove (2) reason inside T and repeat the proof of (1), now using formalized Σ_1 -completeness (see I.4.32). (3) is immediate from the definition of $Proof_\tau^\bullet$. $\square$

2.18 Definition. Let T be a theory containing Q and let π be a definition of the set of all T -provable formulas. The *consistency statement* given by π is the formula $\neg\pi(\overline{0=1})$. We denote the last formula by Con_π or $Con(\pi)$.

2.19 Lemma. Let T and π be as above and assume that π satisfies the provability conditions (1) and (3) (see 2.16). Then, for any T -sentence φ , T proves $Con(\pi) \equiv (\neg\pi(\overline{\varphi}) \vee \neg\pi(\overline{\neg\varphi}))$.

Proof. By (1), $T \vdash \pi(\overline{\neg 0=1})$. Thus

$$\begin{aligned} T \vdash \neg Con(\pi) &\rightarrow \pi(\overline{0=1}) \& \pi(\overline{\neg 0=1}), \\ T \vdash \pi(\overline{0=1} \rightarrow (\neg 0=1 \rightarrow \varphi)) \end{aligned}$$

(by (1)) and hence, by double use of (3), $T \vdash \neg \text{Con}(\pi) \rightarrow \pi(\bar{\varphi})$. Similarly, $T \vdash \neg \text{Con}(\pi) \rightarrow \pi(\neg\bar{\varphi})$.

Conversely, note $T \vdash \pi(\varphi \rightarrow (\neg\varphi \rightarrow 0 = 1))$ by (1), thus double use of (3) gives $T \vdash (\pi(\bar{\varphi}) \& \pi(\neg\bar{\varphi})) \rightarrow \neg \text{Con}(\pi)$. $\square$

2.20 Lemma. Now assume T includes $I\Sigma_1$ and let $Pr_\tau^\bullet$ be as above. If Con is the consistency statement given by $Pr_\tau^\bullet$ then

$$T \vdash \text{Con} \equiv \text{Con}_\tau^\bullet,$$

where $\text{Con}_\tau^\bullet$ is as in I.4.7.

Proof. Fully analogous to the proof of 2.19, but now not as a schema but as a single theorem. $\square$

2.21 Theorem. (Gödel's Second Incompleteness Theorem). Let T be a theory containing Q and let π be a Σ_1 definition of the set of all T -provable formulas satisfying the provability conditions.

Let ν be Gödel's sentence based on π . Then

$$T \vdash \nu \equiv \text{Con}(\pi)$$

and hence T does not prove $\text{Con}(\pi)$ provided T is consistent.

Proof. $T \vdash \nu \rightarrow \neg\pi(\bar{\nu})$, thus $T \vdash \nu \rightarrow \text{Con}(\pi)$. Conversely, $T \vdash \neg\nu \rightarrow \pi(\bar{\nu})$ and, by (2), $T \vdash \pi(\bar{\nu}) \rightarrow \pi(\pi(\bar{\nu}))$, i.e. $T \vdash \pi(\bar{\nu}) \rightarrow \pi(\neg\bar{\nu})$. Therefore we have $T \vdash \neg\nu \rightarrow (\pi(\bar{\nu}) \& \pi(\neg\bar{\nu}))$, i.e. $T \vdash \neg\nu \rightarrow \neg \text{Con}(\pi)$. $\square$

2.22 Corollary. Let T be a theory containing $I\Sigma_1$ and let ν be Gödel's sentence based on $Pr_\tau^\bullet$, where τ is a Σ_1 definition of axioms of T . Then $T \vdash \nu \equiv \text{Con}_\tau^\bullet$; thus T does not prove $\text{Con}_\tau^\bullet$, provided T is consistent.

2.23 Remark. In the terminology of 2.5, Gödel's second incompleteness theorem gives us an example of a formula having a unique fixed point and such that this fixed point is an independent formula (neither provable nor refutable) – assuming that our theory T is Σ_1 -sound.

On the other hand, if $\lambda(x)$ is a flexible formula then for any two different fixed points φ, ψ of λ , T does not prove $\varphi \equiv \psi$ (and it is easy to see from the proof of the diagonal lemma that each formula has infinitely many different fixed points).

2.24 Corollary. Peano arithmetic PA is not finitely axiomatizable.

Proof. If PA were finitely axiomatizable then for some k , PA would be equivalent to $I\Sigma_k$; but $PA \vdash \text{Con}(I\Sigma_k^\bullet)$ and, by 2.22, $I\Sigma_k$ does not prove $\text{Con}(I\Sigma_k^\bullet)$. $\square$

2.25 Löb's Theorem. Let T be as in 2.22, i.e. T contains $I\Sigma_1$, τ is a Σ_1 definition of T . Assume that φ is a sentence such that $T \vdash Pr_\tau^*(\bar{\varphi}) \rightarrow \varphi$. Then $T \vdash \varphi$.

Proof. By our assumption, $(T + \neg\varphi) \vdash \neg Pr_\tau^*(\bar{\varphi})$, thus $(T + \neg\varphi) \vdash Con_{\tau+\neg\varphi}^*$ (where $(\tau + \neg\varphi)(x) \equiv \tau(x) \vee x = \bar{\varphi}$), thus by 2.22, $(T + \neg\varphi)$ is inconsistent. $\square$

2.26 Corollary. Let T contain $I\Sigma_1$ and let κ be a fixed point of the formula $Pr_\tau^*(x)$ (where τ is a Σ_1 definition of T), i.e. $T \vdash \kappa \equiv Pr_\tau^*(\bar{\kappa})$. The formula κ is called Henkin's formula (and says "I am provable"). By Löb's theorem, $T \vdash \kappa$.

2.27 Theorem. For each $n \geq 1$, $I\Sigma_n$ is not axiomatizable by Σ_{n+2} formulas.

Proof. First show that $I\Sigma_n$ is not axiomatizable by Π_{n+1} formulas. If it were we would have a finite set S of Π_{n+1} formulas axiomatizing $I\Sigma_n$ (since the last theory is finitely axiomatizable, see I.2.52, but by I.4.33, $I\Sigma_n$ proves the consistency of the set of all true Π_{n+1} formulas and hence, by "snowing"-snowing, we would have $I\Sigma_n \vdash S^* \subseteq Tr(\Pi_{n+1})$ (S^* is $\{\bar{\varphi}_1, \dots, \bar{\varphi}_n\}^*$ where S is $\{\varphi_1, \dots, \varphi_n\}$). Furthermore, since finite axiomatizability of $I\Sigma_n$ is provable in $I\Sigma_1$, we would get $I\Sigma_1 \vdash Con^*(S^*) \rightarrow Con^*(I\Sigma_n)$ and, by the above, $I\Sigma_n \vdash Con^*(S)$. Thus we would get $I\Sigma_n \vdash Con^*(I\Sigma_n^*)$ which contradicts Gödel's second incompleteness theorem.

Now assume that S is a finite set of Σ_{n+2} formulas axiomatizing $I\Sigma_n$, $S = \{\varphi_1, \dots, \varphi_n\}$, $\varphi_i = (\exists x)\psi_i(x)$, where ψ_i is Π_{n+1} . Let us work in $I\Sigma_n$. We have $Tr_{\Sigma_{n+1}}(S^*)$ and $Con^*(S^*) \rightarrow Con^*(I\Sigma_n^*)$; thus for some $x_1, \dots, x_n$ we have $\bigwedge_i Tr_{\Pi_n}(\bar{\psi}_i(x_i))$ and for $S_1^* = \{\bar{\psi}_1(x_1), \dots, \bar{\psi}_n(x_n)\}$ we have $Con^*(S_1^*) \rightarrow Con^*(S^*)$. But $Con^*(S_1^*)$ follows by I.4.33. We have proved $I\Sigma_n \vdash Con^*(I\Sigma_n^*)$ and have again a contradiction with Gödel's second incompleteness theorem. $\square$

2.28 Discussion. Till now we have worked with arbitrary theories containing Q or $I\Sigma_1$ and either have assumed nothing about the complexity of T as a set of formulas (thus as a set of natural number) or have assumed that T is Σ_1 (has a Σ_1 definition). We shall often need stronger assumptions on T ; we shall now list and discuss four most frequent possibilities.

- (a) T is Σ_1 ;
- (b) T is Δ_1 ;
- (c) T is $I\Sigma_1$ -provably Δ_1 ;
- (d) T is Σ_0^{exp}

Let us stress the fact that we identify a theory with the set of its axioms (not the set of provable formulas). It is easy to see that $(d) \Rightarrow (c) \Rightarrow (b) \Rightarrow (a)$; let us comment on this.

For Σ_0^{exp} formulas cf. I.1.28 and I.2.73; we allow bounded quantifiers of the form $(\forall x \leq y)$, $(\forall x \leq 2^y)$ and similarly for $\exists$. For provably Δ_1 sets cf. I.1.51; by I.1.50 and I.1.52 each provably Δ_1 set is Δ_1 and each Δ_1 set is Σ_1 .

Σ_1 sets are often called recursively enumerable and Δ_1 sets are called recursive (cf. Sect. 0). We also know that each primitive recursive set is $I\Sigma_1$ -provably Δ_1 and have promised to show the converse in Chap. IV by model-theoretic means (cf. I.1.54). Concerning Σ_0^{exp} sets, it is easy to prove that they coincide with $\Sigma_0^{exp}(exp)$ sets and with elementary recursive sets as defined e.g. in Grzegorczyk's book (but we shall not need this fact).

Define two theories T, S to be *deductively equivalent* if they prove the same theorems. (Thus deductively equivalent theories are just two different axiomatizations of a deductively closed set of formulas).

2.29 Craig's Theorem. Each Σ_1 theory T is deductively equivalent to a Σ_0^{exp} theory.

Proof. We show that T is deductively equivalent to a provably Δ_1 theory; the reader may show that the same construction gives in fact a Σ_0^{exp} theory. Let τ be a Σ_1 formula defining the set T of axioms, let τ be $(\exists y)\tau(x, y)$ where τ_0 is Σ_0 . Craig's trick is to replace each axiom φ by a sufficiently long conjunction

$$\varphi \& \varphi \& \dots \varphi$$

$n \text{ times}$

(denoted by $Repeat(n, \varphi)$ if the number of conjuncts is n) which makes it possible to bound the quantifier $(\exists y)$. Thus put

$$\delta(x) \equiv (\exists y \leq x)(\exists u, v \leq x)(x = Repeat^*(v, u) \& \tau_0(u, y))$$

where $Repeat^*$ is defined in the obvious way and is Δ_1 in $I\Sigma_1$, thus $\delta(x)$ is also Δ_1 in $I\Sigma_1$. The set S defined by $\delta(x)$ is an axiom system related to T as follows: $\varphi \in T$ iff, for some n , $Repeat(n, \varphi) \in S$. Clearly, S is deductively equivalent to T . $\square$

2.30 Remarks. (1) Observe that in fact we can prove more: if τ is a Σ_1 -definition of T in $I\Sigma_1$ and δ is as above and defines S then $I\Sigma_1$ proves that the theories τ and δ are deductively^{*} equivalent^{*}.

(2) A formalized version of Craig's theorem reads as follows: $I\Sigma_1$ proves that if T is a Σ_1^* theory^{*} then there is a Δ_1^* theory S such that T and S are deductively^{*} equivalent^{*}. (Proof obvious).

(3) Observe that if δ is Δ_1 in $I\Sigma_1$ and defines T then τ binumerates T in $I\Sigma_1$ (and hence in T). (This is because, by 1.23, if A is any set of natural

numbers and τ is its definition and is Δ_1 in a sound $S \supseteq Q$ then τ binumerates A in S .)

2.31 Further Remarks. (1) We know that each Δ_1 set of natural numbers has a Σ_1 binumeration (and a Π_1 binumeration) in T where T is any consistent Δ_1 theory containing Q . Note that a Δ_1 set may have, and in fact always has different Σ_1 binumerations whose equivalence is not provable in T : let $C \subseteq N$ and let $\gamma(x)$ be any Σ_1 binumeration of C . Let ρ be a Rosser sentence of T , let $(\exists y)\beta(x, y)$ be a Σ_1 definition of refutable formulas; put

$$\gamma_1(x) \equiv \gamma(x) \& (\forall y < x) \neg \beta(\bar{p}, y)$$

$$\gamma_2(x) \equiv \gamma(x) \vee (\exists y < x) \beta(\bar{p}, y).$$

Clearly, $T \vdash \gamma_1(x) \rightarrow \gamma(x) \rightarrow \gamma_2(x)$; we show that both δ_1 and γ_2 binumerate C but T does not prove $\gamma_1(x) \equiv \gamma_2(x)$. Observe that for each k , $T \vdash \gamma_1(\bar{k}) \equiv \gamma(\bar{k}) \equiv \gamma_2(\bar{k})$ since for each k

$$T \vdash \neg(\exists y \leq k) \beta(\bar{p}, k)$$

($\neg\rho$ being unprovable), thus all three formulas binumerate the same set; but $(T + \rho)$ is consistent, i.e. $T + (\exists y)\beta(\bar{p}, y)$ is consistent; and in the last theory we prove $(\exists y)(\gamma_2(y) \& \neg\gamma_1(y))$. Thus T does not prove $\gamma_2(y) \rightarrow \gamma_1(y)$.

(2) Let us show that for each Σ_1 -sound Δ_1 theory T containing Q there is a Δ_1 set A such that no formula δ which is Δ in T_1 binumerates A in T (thus the claim of (1) above cannot be improved to a binumeration which is Δ_1 in T). Let T be given and observe the following:

If T proves $(\forall x)(\delta(x) \equiv \pi(x))$ where $\delta \in \Sigma_1$ and $\pi \in \Pi_1$, then for each $k \in N$, $N \models (\delta(\bar{k}) \equiv \pi(\bar{k}))$ (i.e. $N \models \delta(\bar{k}) \vee \neg\pi(\bar{k})$ and $N \models \neg\delta(\bar{k}) \vee \pi(\bar{k})$ — Σ_1 soundness suffices), thus $N \models (\forall x)(\delta(x) \equiv \pi(x))$. Thus for each $k \in N$, either there is a witness for $\delta(\bar{k})$ or there is a witness for $\neg\pi(\bar{k})$.

Now define A as follows: if n is a T -proof of $(\forall x)(\delta(x) \equiv \pi(x))$ (where $\delta \in \Sigma_1, \pi \in \Pi_1$) then look for witnesses for $\delta(\bar{n})$ and $\neg\pi(\bar{n})$. If you find a witness for $\delta(\bar{n})$, define $n \notin A$; if you find a witness for $\neg\pi(\bar{n})$, define $n \in A$. For remaining n define e.g. $n \in A$.

The reader may check that this is a Δ_1 definition of a set A and that if n, δ, π are as above then A differs from the set defined by δ (and π) at least in n .

(3) By 1.23, if φ is a Σ_1 definition of a set A then φ numerates A in each Σ_1 -sound theory $S \supseteq Q$. Let us show that the assumption of Σ_1 -soundness is essential; in general, φ numerates a superset of A . (For example, let A be the set of axioms of PA , let $\pi(x)$ be the formula $x \in PA^*$ (where PA^* is defined by copying the definition of PA in IS_1) and let $\varphi(x)$ be the formula $\pi(x) \vee (\neg Con^*(PA^*) \& x = \overline{0=1})$). Then π is Σ_1 in PA , numerates PA in PA but numerates $(PA + \overline{0=1})$ in the consistent (but Σ_1 -ill) theory $(PA + \neg Con^*(PA^*))$.

(4) Returning again to Craig's theorem, let us observe the following: If τ is a Σ_1 definition of a theory T and δ is the Δ_1 -in- $I\Sigma_1$ definition of the corresponding Craig theory S then (i) T is deductively equivalent to S , (ii) $I\Sigma_1$ proves that τ is deductively* equivalent* to δ , (iii) δ binumerates S in T (even in $I\Sigma_1$) but (iv) τ may numerate a proper extension of T in T .

(d) Pure Extensions of PA

In this and next subsection we shall deal with extensions of PA having the same language as PA . Here we show that each such theory T is essentially reflexive (2.35) and if T is Δ_1 then it has a Δ_2 definition π such that T proves Con_π^* (2.37). This last result shows some limitations of Gödel's second incompleteness theorem but it should not be overestimated; its main use is in technical proofs.

2.32 Definition. (1) For each theory T , let $T \upharpoonright k$ be the set of all axioms of T less than k .

(2) As above, for each finite set $S = \{\varphi_1, \dots, \varphi_n\}$, let S^* denote $\{\bar{\varphi}_1, \dots, \bar{\varphi}_n\}^*$.

(3) Let T be a theory extending $I\Sigma_1$. T is *reflexive* if for each k , T proves $Con^*((T \upharpoonright k)^*)$ (i.e. for each finite subtheory T_0 of T , T proves the consistency of T_0).

2.33 Definition. (1) A theory T' is a *pure extension* of T if T' extends T and has the same language as T .

(2) Let $T \supseteq I\Sigma_1$. T is *essentially reflexive* if each pure extension of T is reflexive.

2.34 Lemma. Let $T \supseteq I\Sigma_1$. T is essentially reflexive iff for each sentence φ , $T \vdash \varphi \rightarrow Con^*({\varphi}^*)$.

Proof. If T is essentially reflexive and φ is a sentence then $(T + \varphi)$ is reflexive, thus $(T + \varphi) \vdash Con^*({\varphi}^*)$, hence $T \vdash \varphi \rightarrow Con^*({\varphi}^*)$. Conversely, let $T \vdash \varphi \rightarrow Con^*({\varphi}^*)$ for each φ and let T' be a pure extension of T ; let φ be the conjunction of all axioms of $T' \upharpoonright k$. Then the last provability immediately gives $T' \vdash Con^*((T' \upharpoonright k)^*)$. $\square$

2.35 Theorem. PA is essentially reflexive.

Proof. Immediately from I.4.34. $\square$

2.36 Theorem. (1) Let T be any consistent pure extension of PA . Then for any k , T is not axiomatizable using only Σ_k formulas.

(2) For each $k \geq 1$, no consistent pure extension of $I\Sigma_k$ is axiomatizable using only Σ_{k+2} formulas.

Proof. Clearly, (1) follows from (2). The proof of (2) is a variant of the proof of 2.27: Let T be a pure extension of $I\Sigma_k$. First, if S is a finite set of Π_{k+1} formulas such that S proves $I\Sigma_k$ then we get $S \vdash \text{Con}^*(S^*)$ using I.4.34, hence S is inconsistent by Gödel's second incompleteness theorem, thus T is not Π_{k+1} axiomatizable (and (1) follows). The generalization to Σ_{k+2} formulas is as in 2.27, again getting $S \vdash \text{Con}^*(S^*)$, a contradiction. $\square$

2.37 Theorem. Let $T \in \Delta_1$ be a pure extension of PA . Then there is a binumeration τ of T in T such that τ is Δ_2 in T and $T \vdash \text{Con}^*(\tau)$.

Proof. By 1.23, T has a Σ_1 binumeration σ in T and a Π_1 binumeration π in T ; let $\delta(x)$ be $\sigma(x) \ \& \ (\forall y \leq x)(\sigma(y) \equiv \pi(y))$. Evidently, δ is a Δ_2 binumeration of T in T . Let $\tau(x)$ be the formula $\delta(x) \ \& \ \text{Con}^*(\pi \upharpoonright x)$ (where $\pi \upharpoonright x(y)$ is $\pi(y) \ \& \ y \leq x$). Then τ is another Δ_2 binumeration of T in T (since, thanks to reflexivity, for each k we have $T \vdash \tau(\bar{k}) \equiv \delta(\bar{k}) \equiv \sigma(\bar{k}) \equiv \pi(\bar{k})$). We prove $T \vdash \text{Con}^*(\tau)$. Let us work in T . We distinguish two cases. First assume $\text{Con}^*(\pi)$; then evidently $\text{Con}^*(\tau)$ (since τ implies π). Second, assume $\neg \text{Con}^*(\pi)$ and let z be the least number such that $\neg \text{Con}^*(\pi \upharpoonright (z+1))$. Then $\text{Con}^*(\pi \upharpoonright z)$ and for all x , $\tau(x)$ implies $(\pi \upharpoonright z)(x)$. Thus $\text{Con}^*(\tau)$ and the proof is complete. $\square$

2.38 Remark. The reader has observed that τ just describes the largest initial segment of δ which is consistent; thus it is not much surprising that $\text{Con}^*(\tau)$ is provable. Reflexivity guarantees that τ defines (and T -binumerates) T . Observe also that the formula $\text{Con}^*(\tau)$ is Π_2 in T .

(e) Interpretability in Pure Extensions of PA

Our aim is to prove the following two theorems.

2.39 Theorem. Let $T \in \Delta_1$ be a pure extension of PA , let $S \in \Delta_1$ be another theory. Then the following are equivalent:

- (i) S is interpretable in T .
- (ii) S is locally interpretable in T , i.e. each finite subtheory of S is interpretable in T .
- (iii) For each k , $T \vdash \text{Con}^*((S \upharpoonright k)^*)$.
- (iv) There is a binumeration τ of S in T such that $T \vdash \text{Con}^*(\tau)$.

2.40 Theorem. If T and S are as in 2.39 and S is also a pure extension of PA then S is interpretable in T iff S is Π_1 conservative for T .

2.41 Proof of 2.39. We prove (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i). (i) $\Rightarrow$ (ii) is evident. To prove (ii) $\Rightarrow$ (iii) it suffices to show that if S_o is finite and interpretable in T then $T \vdash \text{Con}^*(S_o^*)$. We apply the following

2.42 Lemma. Let S_o, T_o be finite theories, let T_o contain $I\Sigma_1$. If S_o is interpretable in T_o then $T_o \vdash \text{Con}^*(T_o^*) \rightarrow \text{Con}^*(S_o^*)$.

Proof. Let $\text{intp}(d)$ mean that d is a tuple containing T_o -formulas $\chi, \psi_p, \dots, \psi_F, \dots, \psi_c \dots$ of respective arities (given by the language of S_o) as well as T_o -proofs of $(\exists x)\chi(x)$ and of the translations of the axioms of S_o . This is a finite sequence since S_o is finite. Clearly, $\text{intp}(d)$ is Δ_1 ; moreover a little checking shows that it is defined by a formula $\text{intp}^*(x)$ which is Δ_1 in $I\Sigma_1$ (just formalize!). Similarly, let for each S_o -formula φ , $i(\varphi)$ be the translation of φ into a T_o -formula given by the above. Similarly, we have a definition i^* , Δ_1 in $I\Sigma_1$, of i . Using Σ_1 -induction we get $T_o \vdash \text{Pr}_{S_o}^*(x) \rightarrow \text{Pr}_{T_o}^*(i^*(x))$; and similarly, T_o proves the properties of i concerning connectives. Thus $T_o \vdash \text{Con}^*(T_o^*) \rightarrow \text{Con}^*(S_o^*)$, as desired. $\square$

2.43 Proof of 2.39 continued. The lemma gives $T \vdash \text{Con}^*((S \upharpoonright k)^*)$ since if $S_o = S \upharpoonright k$ and S_o is interpretable in T then S_o is interpretable in a finite subtheory T_o of T ; and $T \vdash \text{Con}^*(T_o^*)$ due to reflexivity. This completes the proof of (ii) $\Rightarrow$ (iii).

Now we prove (iii) $\Rightarrow$ (iv). Assume $(\forall k)T \vdash \text{Con}^*((S \upharpoonright k)^*)$ and let σ, π be Σ_1 and Π_1 binumerations of S in $I\Sigma_1$ (thus in T); let $\delta(x)$ be $\sigma(x)$ & $(\forall y \leq x)(\sigma(y) \equiv \pi(y))$ (cf. 2.37). Finally, let $\tau(x)$ be $\delta(x)$ & $\text{Con}^*(\delta \upharpoonright x)$. Then $T \vdash \text{Con}^*(\tau)$ and τ binumerates S in T .

(iv) $\Rightarrow$ (i). We prove the arithmetized completeness theorem appropriately relativized. Let $T \vdash \text{Con}^*(\tau)$ where τ is Σ_m (say), τ binumerates S in T . Then τ is Δ_{m+1} in T and therefore T proves that S has a full low Δ_{m+2} model. In fact, the proof of Low basis theorems gives as formulas $(\chi, \psi_S, \psi_+, \psi_*, \dots)$ defining a low Δ_{m+2} model M of S in T . These formulas can be directly taken to define the desired interpretation i . (To see this prove a small "snowing"-snowing lemma for each S -formula ψ :

$$T \vdash x, \dots, y \in M \rightarrow (\psi^i(x, \dots, y) \equiv M \models \bar{\psi}[x, \dots, y]).$$

This completes the proof of 2.39. $\square$

2.44 Proof of 2.40. First assume that S is interpretable in T ; we prove Π_1 -conservation. Let i be an interpretation of S in T . Thanks to the least number principle we may assume that i is absolute with respect to equality, i.e.

$$T \vdash \chi(x) \& \chi(y) \rightarrow x =^i y \equiv x = y.$$

(Otherwise define a factorization).

The interpretation defines, in T , a model $(M, S^i, +^i, *^i, 0^i, \leq^i)$ of the language of arithmetic (nothing is claimed on the existence of a full satisfaction!). Let us work in T . We show that there is an isomorphic embedding F of the universe onto an initial segment of M . (Cf. IV.1.3.) Let $pism(s)$ (partial isomorphism) mean $Seq(s) \ \& \ (s)_0 = \bar{0}_M \ \& \ (\forall i < lh(s) - 1)((s)_{i+1} = S_M((s)_i))$. Show, by induction, that there are (uniquely determined) partial isomorphisms of arbitrary length; their union is the desired F . Verify by induction, that F preserves $+$ and $*$ (better: F is an isomorphic embedding of the universe into M) and that the range of F is an initial segment of M w.r.t. $\leq_M$. Then, outside T , show for each bounded formula we have

$$T \vdash \varphi(x, \dots) \equiv \varphi^i(F(x), \dots).$$

Thus if $(\forall x)\varphi(x)$ is Π_1 and $S \vdash (\forall x)\varphi(x)$ then $T \vdash ((\forall x)\varphi(x))^i$, which gives $T \vdash (\forall x)\varphi^i(F(x))$, thus $T \vdash (\forall x)\varphi(x)$.

Conversely, let T be Π_1 -conservative for S . Since S is reflexive, we have, for each k , $S \vdash Con^*((S \upharpoonright k)^*)$, thus $T \vdash Con^*((S \upharpoonright k)^*)$, thus S is interpretable in T . $\square$

3. Definable Cuts

Pure extensions of PA , studied in the preceding section, are never finitely axiomatized. In this section we shall investigate mainly finitely axiomatized theories containing $I\Sigma_1$ but possibly having a richer language than the language of arithmetic. These may be e.g. fragments $I\Sigma_n$, $B\Sigma_{n+1}$ ($n > 0$) or ACA_0 , Gödel-Bernays set theory and many others. Besides other things, we prove (in (c)) that a consistent finitely axiomatized sequential theory cannot prove full induction, i.e. there is a formula $\varphi(x)$ such that $T \vdash \varphi(\bar{0}) \ \& \ (\forall x)(\varphi(x) \rightarrow \varphi(x+1))$ but T does not prove $(\forall x)\varphi(x)$. (Here x is a variable ranging over numbers, but φ may contain other variables and symbols of the language of T , e.g. in the case of ACA_0 φ may contain quantified set variables.) It follows then that there is a formula $I(x)$ such that T proves that $I(x)$ defines an initial segment (i.e. T proves $I(\bar{0})$, $(\forall x)(I(x) \rightarrow I(x+1))$, $(\forall x, y)(I(x) \ \& \ y \leq x \rightarrow I(y))$ but T does not prove $(\forall x)I(x)$. (Take $I(x)$ to be $(\forall y \leq x)\varphi(y)$.) We call such an I a proper definable cut. Definable cuts will play an important role in the present and the following section. Here, after proving an important theorem on shortening cuts (subsection (a)) we shall strengthen Gödel's second incompleteness theorem to a theorem saying (roughly) that if $T \supseteq I\Sigma_1$ is consistent and finitely axiomatizable and I is a definable T -cut then T cannot disprove the existence of a T^* -proof^{*} z of contradiction, z belonging to I (subsection (b)). In subsection (c) we shall study an alternative notion of provability, called Herbrand provability (or

direct provability) and show, for reasonable finitely axiomatized theories T , that there is a definable cut I such that T *does* prove that in I there is no Herbrand proof of a contradiction in T . Finally, in subsection (d) we prove a very useful criterion of interpretability of finitely axiomatized theories using definable cuts and Herbrand provability.

(a) Definable Cuts and Their Properties

3.1 Definition. Let T contain Q (the language of T may properly extend the language of arithmetic). A formula $I(x)$ with one free variable (understood as a number variable) is a *definable cut in T* (in short, a T -cut) if T proves $I(0)$, $(\forall x)(I(x) \rightarrow I(x+1))$ and $(\forall x, y)(y < x \ \& \ I(x) \rightarrow I(y))$. I is *proper* if T does not prove $(\forall x)I(x)$. (In the case of ACA_0 , be aware of the fact that I is a formula, possibly with bound set variables; I is not a set variable.)

3.2 Remark. (1) Clearly, if I is a T -cut then, for each natural k , $T \vdash I(\bar{k})$.

(2) If $T \supseteq I\Sigma_n$ then clearly no Σ_n formula (and no Π_n formula) is a proper T -cut. But $I\Sigma_n$ has a proper Σ_{n+1} cut: let $n \geq 1$ and let $(\exists y)\psi(x, y)$ be a Σ_{n+1} formula for which T does not prove induction. Take the formula

$$(\exists s)(Seq(s) \ \& \ lh(s) = x \ \& \ (\forall i < x)\psi(i, (s)_i)).$$

It is an $I\Sigma_n$ -cut and is Σ_{n+1} in $I\Sigma_n$. For $n = 0$ a similar proof works but one has to use a coding of sequences developed in Chap. V.

3.3 Definition. Define 2_y^x in $I\Sigma_1$ by the following recursion: for each x ,

$$2_0^x = x, \quad 2_{y+1}^x = 2^{2_y^x}.$$

Similarly, we define $\omega_y(0) = 0$ and for $x > 0$, $\omega_0(x) = 2x$, $\omega_{y+1}(x) = 2^{\omega_y(|x|-1)}$, where $|x|$ is the upper integral part of the binary logarithm of $x+1$, i.e. $y = |x|$ iff y is the least z such that $2^z > x$.

3.4 Lemma. $I\Sigma_1$ proves the following:

- (1) $(\forall y)(\omega_y(x) \leq 2^x),$
- (2) $(\forall y)2_{y+1}^{x+1} = \omega_y(2_{y+1}^x),$
- (3) $\omega_y(x) \leq \omega_y(x+1),$
- (4) $\omega_y(x) \leq \omega_{y+1}(x),$

Proof elementary. Note that the present use of the symbol ω differs from its use in Chap. II (where we dealt with ordinals). Both uses are common in the literature and we prefer to be in accordance with it. $\square$

3.5 Theorem. Let $T \supseteq I\Sigma_1$. For each $k \geq 0$ and each T -cut I there is a T -cut J such that

$$\begin{aligned} &(\forall x \in J)(2_k^x \in I), \\ &(\forall x \in J)(\omega_k(x) \in J). \end{aligned}$$

Proof. (a) First we prove that for each T -cut I there is a T -cut J such that, in T , $J \subseteq I$ and J is closed under addition: just put

$$J(x) \equiv I(x) \& (\forall y)(I(x+y)).$$

Clearly, J is a cut, and $J \subseteq I$; if $x, z \in J$ then, for each $y \in I$, $z+y \in I$ and therefore $x+z+y \in I$, thus $x+z \in J$. (Consequently, J is closed under ω_0 .)

(b) For each n , there is a T -cut J_n such that $T \vdash J_n \subseteq T$ and $T \vdash (\forall x)(x \in J_n \rightarrow 2_n^x \in I)$. This is proved by induction on n : Assume J_n given. By shortening we may assume that J_n is closed under addition. Define $x \in J_{n+1} \equiv 2^x \in J_n$. Then in T , $x \in J_{n+1} \rightarrow 2^x \in J_n \rightarrow (2)_n^{2^x} \in I$ and J_{n+1} is a cut: if $x \in J_{n+1}$ then $2^{x+1} = 2^x + 2^x \in J_n$ thus $x+1 \in J_{n+1}$.

(c) For each n , there is a T -cut K_n such that $T \vdash K_n \subseteq I$ and $T \vdash (\forall x)(x \in K_n \rightarrow \omega_n(x) \in K_n)$ (i.e. K_n is closed under ω_n). First let J_n be as in (c) and put

$$x \in K_n \equiv (\exists y \in J_n)(x \leq 2_n^y)$$

Then clearly $T \vdash J_n \subseteq K_n \subseteq I$; in T , if $x \in K_n$ and $x \leq 2_n^y$ for some $y \in J_n$ then $\omega_n(x) \leq \omega_n(2_n^y) = 2_n^{y+1}$ and $y+1 \in J_n$, thus $\omega_n(x) \in K_n$. Finally, K_n is a cut: in T , if $x \in K_n$ and $x \leq 2_n^y$ for $y \in J_n$ then $x+1 \leq 2_n^{y+1}$ (for each n , $I\Sigma_1$ proves $2_n^{y+1} \geq 2_n^y + 1$ – show by induction on n).

(d) Theorem 3.5 follows by (b) and (c). $\square$

(b) A Strong Form of Gödel's Second Incompleteness Theorem

We shall investigate existence of proofs of a contradiction in a definable cut. Theorem 3.9 is a general theorem for $T \supseteq Q$ (not mentioning cuts), 3.11 is a consequence for $T \supseteq I\Sigma_1$ saying that, if consistent, T does not prove the consistency of T over a cut. Note that using the apparatus of Chap. V, 3.11 can be generalized for $T \supseteq Q$ (but a careful formulation is necessary, see V.5.28 (ii)). The key device for 3.9 is simultaneous use of two provability predicates.

3.6. Let $T \supseteq Q$ and let π_1, π_2 be two definitions of T -provable formulas. Generalized provability conditions for π_1 and π_2 over T are the following

three conditions: For all φ ,

- (1) $T \vdash \varphi$ implies $T \vdash \pi_1(\overline{\varphi})$,
- (2) $T \vdash \pi_1(\overline{\varphi}) \rightarrow \pi_2(\overline{\pi_1(\overline{\varphi})})$,
- (3) $T \vdash \pi_1(\overline{\varphi}) \& \pi_1(\overline{\varphi \rightarrow \psi}) \rightarrow \pi_2(\overline{\psi})$.

3.7 Remark. Observe that (1) and (3) imply $T \vdash \pi_1(\overline{\varphi}) \rightarrow \pi_2(\overline{\varphi})$: take φ for ψ .

3.8 Definition. The consistency statement for π based on φ is the formula $\neg\pi(\overline{\varphi}) \vee \neg\pi(\overline{\neg\varphi})$. We may denote the last formula by $Conb(\pi, \varphi)$ (cf. 2.18).

3.9 Generalization of Gödel's Second Incompleteness Theorem. Let $T \supseteq Q$ be consistent and let π_1, π_2 satisfy the generalized provability conditions over T . Then there is a sentence φ such that T does not prove the consistency statement $Conb(\pi_2, \varphi)$.

Proof. Let ν be a fixed point of $\neg\pi_1(x)$, i.e. $T \vdash \nu \equiv \neg\pi_1(\overline{\nu})$. This is like Gödel's fixed point but now π_1 need not be Σ_1 and need not define T -provable formulas. Nevertheless, condition (1) is sufficient to show that T does not prove ν (cf. 2.8). Let φ be $\pi_1(\nu)$. We show $T \vdash \neg\pi_2(\overline{\varphi}) \rightarrow \nu$ and $T \vdash \neg\pi_2(\overline{\neg\varphi}) \rightarrow \nu$; thus T does not prove $\neg\pi_2(\overline{\varphi}) \vee \neg\pi_2(\overline{\neg\varphi})$.

First we prove $T \vdash \neg\nu \rightarrow \pi_2(\overline{\pi_1(\overline{\nu})})$. This is because $T \vdash \neg\nu \rightarrow \pi_1(\nu)$ and $T \vdash \pi_1(\nu) \rightarrow \pi_1(\overline{\nu})$ (by (2)).

Second, we prove $T \vdash \neg\nu \rightarrow \pi_2(\overline{\neg\pi_1(\overline{\nu})})$. Indeed, $T \vdash \nu \rightarrow \neg\pi_1(\overline{\nu})$, thus $T \vdash \pi_1(\nu \rightarrow \neg\pi_1(\overline{\nu}))$ by (1); furthermore, $T \vdash \neg\nu \rightarrow \pi_1(\overline{\nu})$, which, together with the preceding provability gives $T \vdash \neg\nu \rightarrow \pi_2(\overline{\neg\pi_1(\overline{\nu})})$ by (3). This completes the proof. $\square$

3.10 Definition. Let $T \supseteq I\Sigma_1$, let I be a T -cut and let τ be a Σ_0^{exp} definition of T . Then $Pr_\tau^{\bullet I}(x)$ is the formula

$$(\exists z)(I(z) \& Proof_\tau^\bullet(z, x))$$

(saying that there is a τ -proof $^\bullet$ of x in I).

(2) $Con_\tau^{\bullet I}$ is the formula $\neg(\exists z)(I(z) \& Proof_\tau^\bullet(z, \overline{0=1}))$.

3.11 Another Generalization of Gödel's Second Incompleteness Theorem. Let $T \supseteq I\Sigma_1$, let I be a T -cut and τ a Σ_0^{exp} -definition of T . Then T does not prove $Con_\tau^{\bullet I}$.

3.12 Remark. The rest of this subsection contains a proof of 3.11. Without any loss of generality we shall assume that T contains a unary function symbol

for exponentiation (2^x) and the corresponding axioms. (It is easy to check that all uses of 2^x may be understood as abbreviations.)

3.13 Definition and Discussion. A function F of one argument is multi-exponentially bounded in $T \supseteq I\Sigma_1$ if there is a k such that $T \vdash (\forall x)(F(x) < 2_k^x)$.

Let us indicate that natural functions describing syntax (as concatenation of two sequences, concatenation of a sequence of sequences, substitution etc.) are multi-exponentially bounded in $I\Sigma_1$. To simplify matters, we shall just show existence of multiexponential bounds in the standard model N ; the formalization is obvious.

(1) Call a function $F(m, \dots, n)$ of several arguments multi-exponentially bounded if there is a k since that, for each q ,

$$(m < q \ \& \dots \ \& \ n < q) \rightarrow F(m, \dots, n) < 2_k^q.$$

Observe that each term of the language $L_0(exp)$ (i.e. having the constant $\bar{0}$ and function symbols $S, +, *, 2^x$) defines a monotone and multi-exponentially bounded function. Write m.e.b. for "multi-exponentially bounded".

(2) Recall that the pairing function is polynomially bounded: $(x, y) \leq (x + y + 1)^2$. Furthermore, recall that our coding of finite sets of numbers (introduced in $I\Sigma_1$) satisfies $x \subseteq (< y) \rightarrow x < 2^y$ and $x \subseteq (< y) \times (< 2) \rightarrow x < 2^{(y+2+1)^2}$. This implies that if s is a sequence $(Seq(s))$, $lh(s) \leq x$ and each member of s is $< y$ then $s \leq 2^{(x+y+1)^2}$. If s, t are two such sequences then their concatenation $s \frown t$ satisfies $s \frown t \leq 2^{(2x+y+1)^2}$; if q is a sequence of sequences, $lh(q) = z$ and each member of q is a sequence satisfying the assumptions above then the concatenation of q (denoted $Concseq(q)$ in Chap. I) satisfies $Concseq(q) < 2^{(2x+y+1)^2}$. The concrete formulas are not too important; what is important is the fact that the bounds are multi-exponential. Now it is easy to see that the substitution function $Subst(\varphi, x, \tau)$ is m.e.b. and similarly for other syntactic notions.

3.14 Lemma. Let T and τ be as in 3.11, i.e. $T \subseteq I\Sigma_1$, τ a Σ_0^{exp} -definition of T .

(a) If φ is Σ_0^{exp} then there is a k such that

$$\begin{aligned} T \vdash x < u \ \& \dots \ \& \ y < u. \\ \rightarrow [\varphi(x, \dots, y) \rightarrow (\exists z < 2_k^u) \text{Proof}_\tau^\bullet(z, \bar{\varphi}(\dot{x}, \dots, \dot{y}))]. \end{aligned}$$

(b) Similarly, if I is a T -cut then there is a k such that

$$T \vdash I(x) \rightarrow (\exists z < 2_k^x) \text{Proof}_\tau^\bullet(z, \bar{I}(\dot{x})).$$

Proof. (a) The proof is a re-examination and generalization of the proof of Σ_1 -completeness of Q (I.1.8). For simplicity prove again a non-formalized version saying the following:

if $m < p \& \dots \& n < p$ and $N \models \varphi(\overline{m}, \dots, \overline{n})$ then there is a T -proof of $\varphi(\overline{m}, \dots, \overline{n})$ beneath 2_k^p (where k is a constant depending only on φ).

(1) First assume that φ is $\overline{m} + \overline{n} = \overline{m + n}$. The usual proof of this formula looks as follows:

$$\begin{aligned} x + \overline{0} &= x \\ x + Sy &= S(x + y) \\ \overline{m} + \overline{0} &= \overline{m} \\ \overline{m} + S\overline{0} &= S(\overline{m} + \overline{0}) \\ \overline{m} + S\overline{0} &= \overline{m + 1} \\ &\dots \\ \overline{m} + \overline{n} &= S(\overline{m} + \overline{n - 1}) \\ \overline{m} + \overline{n} &= \overline{m + n} \end{aligned}$$

It depends on details of your Hilbert-style formalism whether this is a proof as it stands or if you have to make some inessential modifications; but in any case, the length of the proof (i.e. the number of proof lines) is polynomial (here linear) in $\max(m, n)$ and the length of each row is polynomial (linear) in $\max(m, n)$. Thus, by the above the whole proof is m.e.b. in $\max(m, n)$.

Similarly for $\overline{m} * \overline{n} = \overline{m * n}$, $\overline{m} \neq \overline{n}$ (if $m \neq n$), and other cases (cf. I.1.8). Also the proof of $\overline{2^m} = 2^{\overline{m}}$ is easily estimated.

(2) Let $Val(t(m, \dots, n))$ be the value of a closed term (possibly containing exponentiation). For a given t there is a k such that for $m, \dots, n < q$ there is a $d < 2_k^q$ such that d is a T -proof of $t(\overline{m}, \dots, \overline{n}) = \overline{Val(t(m, \dots, n))}$.

To see this, first note that, by (1) in the proof of 3.13, $Val(t(\overline{m}, \dots, \overline{n}))$ is m.e.b., i.e., for some h , whenever $m, \dots, n < q$ then $Val(t(\overline{m}, \dots, \overline{n})) < 2_h^q$. Put $r = 2_h^q$. For simplicity, just take one example: let $t(\overline{m}, \overline{p}, \overline{n})$ be $(2^{\overline{m}} + \overline{n}) * \overline{p}$; we want to estimate a proof of $(2^{\overline{m}} + \overline{n}) * \overline{p} = \overline{(2^m + n) * p}$. By (1) here, we can successively produce proofs of

$$\begin{aligned} 2^{\overline{m}} &= \overline{2^m}, \\ \overline{2^m} + \overline{n} &= \overline{2^m + n}, \\ (\overline{2^m + n}) * \overline{p} &= \overline{(2^m + n) * p}. \end{aligned}$$

There is a common j such that there is a proof of each of these equalities beneath 2_j^r (since each argument involved is $\leq r$). A proof of the desired equality results by concatenating proofs of the equalities above and adding some few lines (instances of transitivity of $=$). Clearly, the whole proof is a m.e.b. function of r , i.e. of the initial arguments $m, \dots, n$.

(3) It follows that if $\varphi(x, \dots, y)$ is a true atomic or negated atomic formula then the function assigning to $m, \dots, n$ the least T -proof of $\varphi(\bar{m}, \dots, \bar{n})$ is m.e.b.

(4) Prove by induction the following: for each Σ_0^{exp} -formula $\varphi(x, \dots, y)$, the function assigning to each $m, \dots, n$ such that $\varphi(\bar{m}, \dots, \bar{n})$ is true, a T -proof of $\varphi(\bar{m}, \dots, \bar{n})$, and to each $m, \dots, n$ such that $\varphi(\bar{m}, \dots, \bar{n})$ is false, a T -proof of $\neg\varphi(\bar{m}, \dots, \bar{n})$ is m.e.b. This is true for φ atomic (see (3)); the induction step for connectives is easy. It remains to handle bounded quantifiers $(\forall x \leq y)$ and $(\forall x \leq 2^y)$.

(5) Observe that T proves

$$(\forall x \leq \bar{m})\varphi(x, \bar{m}, \dots, \bar{n}) \equiv \bigwedge_{j \leq m} \varphi(\bar{j}, \bar{m}, \dots, \bar{n})$$

and we can find a proof of this equivalence by m.e.b. function (with arguments $m, \dots, n$). Indeed, analysing I.1.6 (4) we see that finding a proof of

$$x \leq \bar{m} \equiv x = \bar{0} \vee \dots \vee x = \bar{m}$$

is m.e.b. (in the argument m) and so are functions witnessing the following T -provabilities:

$$\begin{aligned} & (\forall x \leq \bar{m})\varphi(x, \bar{m}, \dots) \rightarrow \varphi(\bar{k}, \bar{m}, \dots), \quad (k \leq m) \\ & (\forall x \leq \bar{m})\varphi(x, \bar{m}, \dots) \rightarrow \bigwedge_{k \leq m} \varphi(\bar{k}, \bar{m}, \dots), \\ & \varphi(\bar{k}, \bar{m}, \dots) \rightarrow (x = k \rightarrow \varphi(x, \bar{m}, \dots)), \\ & \bigwedge_{k \leq m} \varphi(\bar{k}, \bar{m}, \dots) \rightarrow \bigvee_{k \leq m} x = b. \rightarrow \varphi(x, \bar{m}, \dots), \\ & \bigwedge_{k \leq m} \varphi(\bar{k}, \bar{m}, \dots) \rightarrow (x \leq \bar{m} \rightarrow \varphi(x, \bar{m}, \dots)), \\ & \bigwedge_{b \leq m} \varphi(\bar{k}, \bar{m}, \dots) \rightarrow (\forall x \leq \bar{m})\varphi(x, \bar{m}, \dots). \end{aligned}$$

Similarly for $(\forall x \leq \bar{2}^m)$, i.e. for $(\forall x \leq 2^{\bar{m}})$: provably a function m.e.b. in the argument 2^m (or even 2_k^m) is m.e.b. in m . This completes the proof of 3.14 (a).

Now we prove (b); this is much easier.

Assume that $I(x)$ is a T -cut; let d be a T -proof of $I(\bar{0}) \ \& \ (\forall x)(I(x) \rightarrow I(x+1))$. Let m be given; a proof of $I(\bar{m})$ consists of proofs of $I(\bar{0}), I(\bar{1}), \dots, I(\bar{m})$:

$$\begin{aligned}
& \dots \\
& I(\bar{0}) \\
& I(x) \rightarrow I(x+1) \\
& I(\bar{0}) \rightarrow I(\bar{1}) \\
& I(\bar{1}) \\
& \dots \\
& I(\overline{m-1}) \rightarrow I(\overline{m}) \\
& I(\overline{m}).
\end{aligned}$$

It is easily seen that this is a m.e.b. function of m , I , d , i.e. for fixed I and d , a m.e.b. function of m . This completes the proof of 3.14, except for the following pedantical

Remark: In proving (a) we in fact assumed $T \supseteq I\Sigma_1(exp)$, i.e. used explicitly the power-of-two operation. If $T \supseteq I\Sigma_1$ but does not have 2^x in the language then we may extend it by adding the definition of 2^x ; this is a particular conservative extension and one can check that there is a m.e.b. function F assigning to each $T(exp)$ -proof of a T -formula φ a T -proof of φ . This shows how to get rid of 2^x . $\square$

3.15 Lemma. Let T , τ be as above. For each T -cut I there is a T -cut J such that $T \vdash J \subseteq I$ and Pr_τ^J , Pr_τ^I satisfy the provability conditions 3.6.

Proof. Let us write Pr^I instead of Pr_τ^I and similarly for Pr^J . The condition (1) is evidently satisfied: if d is a T -proof of φ then $T \vdash Proof_\tau^*(\bar{d}, \bar{\varphi})$ and $T \vdash I(\bar{d})$. To get (3) it suffices to have $T \vdash J(x) \rightarrow I(2_n^x)$ where, provably in T , $u, v, w \subseteq x \rightarrow u \cap v \cap w \leq 2_n^x$. We prove (2). Let us work in T .

Assume $Pr^J(\bar{\varphi})$, i.e., $J(x) \ \& \ Proof^*(\bar{\varphi}, x)$. Then, by 3.14, for an appropriate j given from outside there are $z_1, z_2 < 2_j^x$ such that $Proof^*(\bar{J}(\bar{x}), z_1)$ and $Proof^*(\overline{Proof^*(\bar{\varphi}, \bar{x})}, z_2)$. Then $z = z_1 \cap z_2 \cap \langle Pr^J(\bar{\varphi}) \rangle$ (or something very similar, details are unimportant) is a τ -proof^o of $Pr^J(\bar{\varphi})$; thus it suffices that z is in I .

We see that for example putting $k = h + j$ and choosing J such that $T \vdash J(x) \rightarrow I(2_k^x)$ (by 3.5), we have (1), (2), (3). This completes the proof. $\square$

3.16 Proof of 3.11-conclusion. Let T , τ , I be as in 3.11 and let J_0 be the cut constructed in 3.15. Then for some particular φ T does not prove $Conb(\tau, \varphi)$. To get the unprovability of $\bar{0} = \bar{1}$ in the desired cut J , we shorten J_0 again in such a way that T proves the following: if $x \in J$ is a τ -proof of φ and $y \in J$ is a τ -proof of $\neg\varphi$ then the concatenation of x, y , and the (standardly long) propositional proof of $\bar{0} = \bar{1}$ from φ and $\neg\varphi$ is in J_0 (use e.g. the h above). This completes the proof of 3.11. $\square$

(c) Herbrand Provability and Herbrand Consistency

In this subsection we are going to investigate an alternative notion of provability called Herbrand provability (since it is based on Herbrand's theorem). Even if Herbrand provability is equivalent to the usual provability (provably in $I\Sigma_1$), Herbrand proofs are much "slower" (or: longer) than the usual Hilbert style proofs; we shall show that under some conditions on a theory T , we may always find a cut I in T such that T proves that in I there is no Herbrand proof of inconsistency in T . This is interesting at its own; but the methods we present here will be very useful in studying interpretability and partial conservativity in the next section.

Recall Herbrand's theorem (0.21 and, formalized, 1.4.15): it says that a formula Φ is provable (in predicate calculus) iff there is a disjunction D of instances of the open part of the Herbrand form $He(\Phi)$ of Φ such that Φ is a propositional tautology. This leads to the following

3.17 Definition ($I\Sigma_1$). A *H-proof* (Herbrand proof) of a formula x is a propositional proof z of a disjunction of instances of the open part of $He^*(x)$. (Notation: $HProof^*(z, x)$.) If T is a finite theory consisting of closed formulas then a *H-proof* of x in T is a *H-proof* z of $(\bigwedge T \rightarrow x)$ ($\bigwedge$ standing for a finite conjunction); notation $HProof_T^*(z, x)$. A formula x is *H-provable* in T ($HPr_T^*(x)$) if there is a *H-proof* of x in T . T is *H-consistent* ($HCon^*(T)$) if there is no *H-proof* of $\neg \bigwedge T$.

3.18 Remark. We know that $I\Sigma_1$ proves Herbrand's theorem, i.e. $I\Sigma_1 \vdash HPr_T^*(x) \equiv Pr_T^*(x)$ and $I\Sigma_1 \vdash HCon^*(T) \equiv Con^*(T)$. But for weaker theories not proving Herbrand's theorem, these notions may differ. This will be discussed in Chap. V; here we pay attention to the fact that the notions of provability and *H-provability* may differ on a definable cut, i.e. there may be a cut provably not containing a *H-proof* of a given formula (expressing consistency) but the same cannot be proved for usual Hilbert-style proofs.

3.19 Definition. Let T be a finite theory containing $I\Sigma_1$ (i.e. we assume that T proves all the axioms of $I\Sigma_1$) and let I be a T -cut. We make the following definitions:

$$\begin{aligned} HPr_T^{*I}(x) &\equiv (\exists y)(I(y) \& HProof_T^*(x, y)), \\ HCon_T^{*I} &\equiv \neg HPr_T^{*I}(\neg \bigwedge T^*). \end{aligned}$$

We also write $HCon^{*I}(T^*)$.

3.20 Theorem. Let T be a finite theory in the language of arithmetic extending $I\Sigma_1$. (1) There is a T -cut I such that

$$T \vdash HCon^{*I}(T^*).$$

(2) More generally, there is a T -cut I such that T proves

$$(\forall u \Sigma_1\text{-sentence}^*)(Tr_{\Sigma,1}(u) \rightarrow HCon^{\bullet I}(T^{\bullet} + u)).$$

3.21 Theorem. Still more generally, let T be a finitely axiomatized sequential theory containing IS_1 (i.e. an interpretation of IS_1 in T has been fixed). Then there is a T -definable cut I such that T proves

$$(\forall u \Sigma_1\text{-sentence}^*)(Tr_{\Sigma,1}(u) \rightarrow HCon^{\bullet I}(T^{\bullet} + u)).$$

Let us make a definition and describe the method of proof; then we shall elaborate the proof in details.

3.22 Definition. For each k we define Q_k -formulas as certain formulas of the language of arithmetic. Q_0 -formulas are Σ_0 -formulas; Q_{k+1} -formulas are Q_k -formulas and boolean combinations of formulas of the form $(\exists x)\varphi$ where φ is Q_k (say, boolean combinations of $\exists Q_k$ -formulas). Thus a Q_k -formula has nesting of quantifiers at most k , Σ_0 -formulas being disregarded. A Q_k -proof is a (Hilbert style) proof consisting only of Q_k -formulas.

3.23 Remark. Even if a detailed proof of our theorem is rather long, the idea is easily comprehensible. We shall prove 3.20 (2); then we indicate how to get 3.21. First we show that for each k , satisfaction and truth for Q_k -formulas is definable in IS_1 . It follows that there is a T -cut J such that Q_k -proofs lying in J preserve truth. Consequently, no Q_k -proof of $\neg \bigwedge T^{\bullet}$ is in J . Finally, it can be shown that each H -proof of a Q_k -formula can be transformed, in a multi-exponentially bounded way, to a Q_{k+2} -proof of the same formula. Thus if I is a suitable shortening of J then I has the desired properties.

3.24 Corollary. Let $T \supseteq IS_1$ be consistent, finite and sequential. Then T does not prove full induction.

Proof. If T proves full induction then for each T -cut I , T proves $(\forall x)I(x)$, i.e. there are no proper T -cuts. On the other hand, if T is as above, then $T \vdash HCon^{\bullet I}(T^{\bullet})$ for some T -cut I , thus $T \vdash HCon^{\bullet}(T^{\bullet})$. But since $T \supseteq IS_1$, T proves Herbrand's theorem and hence $T \vdash Con^{\bullet}(T^{\bullet})$, which contradicts Gödel's second incompleteness theorem. $\square$

The rest of the subsection elaborates 3.23.

*

3.25 Lemma. For each k , the following can be constructed in IS_1 : a definable cut I_k and a satisfaction for Q_k -formulas from I_k .

Proof. We know that there is a Δ_1 satisfaction for Σ_0 -formulas (see I.1.75); thus let I_0 be an improper cut. Assume $Sat_{Q,k}$ is a satisfaction for Q_k -formulas from I_k (i.e. Tarski's conditions are provable and, consequently, "it's-snowing"-it's-snowing lemma is provable). In the usual way we get the satisfaction $Sat'(z, e)$ for $\exists Q_k$ -formulas from I_k . It remains to extend the satisfaction to boolean combinations of such formulas, but this is easy. Define $Sat_{Q,k+1}(z, e)$, where z is a boolean combination of $\exists Q_k$ -formulas, iff there exists a "constructing sequence" c of z from $\exists Q_k$ -formulas (each member is either a $\exists Q_k$ -formula or results from some preceding members using connectives; the last member is z) and a corresponding "evaluating sequence" s of zeros and ones such that, if $(c)_i$ is $\exists Q_k$ then $(s)_i = 1$ iff $Sat'(z, e)$, s obeys truth tables and the last member of c is 1. Define $I_{k+1}(x)$ iff $I_k(x)$ and for each Q_{k+1} -formula $z \leq y$ there exists a constructing sequence c with a unique evaluating sequence s . Clearly, I_{k+1} is an $I\Sigma_1$ -cut. $\square$

3.26 Lemma. For each k , there is an $I\Sigma_1$ -cut $I(x)$ that is a shortening of I_k and such that $I\Sigma_1$ proves the following: If T is a finite theory in the language of arithmetic whose axioms are true Q_k -formulas*, if s is a Q_k -proof* from T and $I(d)$ then each member of d is true.

Proof. This is trivial: let $I(x)$ say that $I_k(x)$ and each Q_k -proof from true Q_k -formulas is truth-preserving. This is an inductive property since deduction rules are truth preserving and logical axioms are true. $\square$

3.27 Remark. If $T \supseteq I\Sigma_1$ is a finite theory in the language of arithmetic then for some k , all axioms of T are Q_k and T proves all its axioms true (by "it's-snowing"-it's-snowing). Thus it follows that in the cut I from 3.26 there is no Q_k -proof of $\neg \bigwedge T$. More than that: if u is Σ_1^* and true then it is a true Q_k formula (since $\Sigma_1 \subseteq Q_1$ and thus I does not contain any Q_k -proof of $\neg \bigwedge T \vee u$). To complete the proof of 3.20 it suffices to clarify the relation of H -proofs to Q_k -proofs. Before doing that let us introduce a technical device, analogous to a Henkin extension.

3.28 Definition ($I\Sigma_1$). Take the predicate calculus with the language of arithmetic; using the method of I.4, add infinitely many constants in such a way that for each formula $\varphi(x)$ of the enriched language with just one variable we have a constant $c_{(\forall x)\varphi(x)}$ (associated to φ by a total Δ_1 function) and call the formula

$$\varphi(c_{(\forall x)\varphi(x)}) \rightarrow (\forall x)\varphi(x)$$

the *special axiom* for $\varphi(x)$.

3.29 Remark. (1) Analogously to I.4.9 one shows that this is a conservative extension of the predicate calculus, see also below.

(2) The following theorem is in fact a strengthening of one implication from Herbrand's theorem; namely, it shows that Herbrand provability implies provability. But we need more: Q_k -provability and multi-exponential growth.

3.30 Theorem. For each $k \geq 1$, there is a q such that $I\Sigma_1$ proves the following: There is a Δ_1 function $F(d)$, majorized by 2_q^d , such that if d is a H -proof^{*} of a Q_k -formula^{*} φ then $F(d)$ is a Q_{k+2} -proof^{*} of φ .

The rest of this subsection contains a proof. We shall carefully describe the construction of the proof $F(d)$ and check that it is a Q_{k+2} -proof^{*}; the tiresome task of checking that F is m.e.b. is largely left to the reader. The proof is an inspection and elaboration of a proof from Shoenfield's book. (Alternatively, the reader may apply Theorem V.5.14)

3.31. First, let us analyse a little bit the construction of a prenex normal form of a given formula (such a construction is the initial part of the construction of the Herbrand variant, cf. 0.18). Let Φ be given; we may assume that negation occurs only before atomic formulas, distinct quantifiers bind distinct variables and no variable is both free and bound in Φ . (If not then the corresponding changes as well as the proof of equivalence of both forms are given by a m.e.b. Δ_1 function.) Call, for a moment, a formula *clean* if it satisfies our assumptions. A clean formula is a boolean combination (using $\&$, $\vee$, $\rightarrow$) of atoms, negated atoms and quantifies formulas, say $\beta(\varphi_1, \dots, \varphi_n)$. If φ_1 is $(\Box x)\alpha$ (where $\Box$ is $\forall$ or $\exists$) then $\beta(\varphi_1, \dots, \varphi_n)$ is provably equivalent to $(\Box' x)\beta(\alpha, \varphi_2, \dots, \varphi_n)$ where $\Box'$ is $\forall$ or $\exists$; there is a m.e.b. function finding $\Box'$ and giving a proof of the equivalence. Call the transition from $\beta(\varphi_1, \dots, \varphi_n)$ to $(\Box' x)\beta(\alpha, \dots, \varphi_n)$ the *extraction* of $\Box$ from $(\Box x)\alpha$ in $\beta(\varphi_1, \dots, \varphi_n)$, the inverse transition is the *insertion* of $\Box'$ into $\beta(\alpha, \dots, \varphi_n)$ by quantifying α . A prenex normal form (*PNF*) of Φ is constructed by a sequence of extractions of quantifiers:

$$\Phi = \Phi_0 = (\Box_0 \dots) \Psi_0 \quad (\text{where } \Box_0 \text{ is an empty block of quantifiers})$$

...

$$\Phi_h = (\Box_h \dots) \Psi_h \quad (\Box_h \text{ is a block of quantifiers and } \Psi_h \text{ is open});$$

here for each $i < k$, $(\Box_{i+1} \dots)$ is either of the form $(\Box_i \dots)(\forall x)$ or of the form $(\Box_i \dots)(\exists x)$ and the formula $(\forall x)\Phi_{i+1}$ (or $(\exists x)\Psi_{i+1}$) results from Ψ_i by an extraction of a quantifier. Φ_h is a *PNF* of Φ ; write $(\Box \dots)K$ instead of $(\Box \dots)\Psi_h$.

3.32. The Herbrand normal form $He(\Phi)$ can be written as $(\exists \dots)K^*$ where $(\exists \dots)$ is the block of existential quantifiers resulting from $(\Box \dots)$ in $(\Box \dots)K$ by deleting universal quantifiers and K^* results from K by substitution of certain terms (containing new function symbols) for variables that were universally quantified in $(\Box \dots)K$ (call these variables $\forall$ -variables and the

others $\exists$ -variables). Now, by 3.17, a Herbrand proof of Φ is a propositional proof of a disjunction $\bigvee_i K(t_i)$, where each t_i is a tuple of terms and $K(t_i)$ is in fact an instance of K^* . We may assume that the disjunction $\bigvee_i K(t_i)$ contains no variables.

3.33. We shall successively eliminate new function symbols, replacing them by special constants. Write t_i^0 for t_i ; we shall construct a sequence of propositional tautologies

$$\begin{aligned} & \bigvee_i K(t_i^0), \\ & \quad \dots \\ & \bigvee_i K(t_i^p); \end{aligned}$$

each $K(t_i^j)$ will be an instance of K (but not necessarily of K^*). The transition from j to $j+1$ proceeds as follows:

Take a term $f(a)$ occurring in $\bigvee_i K(t_i^j)$ where f is a Herbrand function (not occurring in Φ) and the terms a contain no Herbrand functions. Replace $f(a)$ (in all occurrences in $\bigvee_i K(t_i^j)$) by the special constant c defined as follows:

The function f corresponds to a quantifier $(\forall x)$ in $PNF(\Phi)$ that was extracted when going from $(\Box_u \dots) \Psi_u$ to $(\Box_u \dots)(\forall x) \Psi_{u+1}(x, y)$; c is the special constant for $(\forall x) \Psi_{u+1}(x, a)$, i.e. the corresponding special axiom is

$$\Psi_{u+1}(c, a) \rightarrow (\forall x) \Psi_{u+1}(x, a).$$

This change clearly preserves the property of being a tautology (since it commutes with all connectives). We finally arrive at a tautology $\bigvee_i K(t_i^p)$ which does not contain any new function. Its propositional proof is obtained from the original propositional proof by a m.e.b. function.

3.34. We want to show that each disjunct $K(t_i^p)$ implies our original formula Φ by a Q_{k+2} -proof using possibly special axioms for constants used in the construction. Write $K(s)$ for $K(t_i^p)$. Recall that $K(s)$ is in fact $\Psi_h(s)$ (see the construction of $PNF(\Phi)$). Now Φ_h is $(\Box_h \dots) \Psi_h$ and

$$(\Box_h) \Psi_h = \begin{cases} (\Box_{h-1})(\forall x) \Psi_{h-1}(x, y) & \text{(Case 1)} \\ (\Box_{h-1})(\exists x) \Psi_{h-1}(x, y) & \text{(Case 2)} \end{cases}$$

where $\Psi_{h-1}(y)$ results from $(\Box_x) \Psi_h(x, y)$ by inserting $(\Box_x)$ in its place.

Case 1. The variable x corresponds in s to a special constant c belonging to the special axiom $\Psi_h(c, s') \rightarrow (\forall x) \Psi_h(x, s')$ (where s' is the rest of s , i.e. s is (c, s')). Denote this axiom by $Sp(c)$. Thus we have a proof of

$$Sp(c) \& \Psi_h(c, s') \rightarrow \Psi_{h-1}(s')$$

and the proof is a Q_{k+1} -proof (since each Ψ_i is a Q_k -formula; Sp has one more unbounded quantifier and a proof of the equivalence of $(\forall x)\Psi_h(x, s')$ and $\Psi_{h-1}(s')$ uses only Q_{k+1} -formulas).

Case 2. Here let s be (d, s') where d is some term corresponding to x ; clearly, the formula $\Psi_h(d, s') \rightarrow (\exists x)\Psi_h(x, s')$ has a Q_{k+1} -proof and the same holds for $\Psi_h(d, s') \rightarrow \Psi_{h-1}(s')$.

3.35. Iterating this we finally construct a Q_{k+1} -proof of

$$\bigwedge_j Sp(c_j) \& K(s) \rightarrow \Phi$$

(since Φ is Ψ_0). Since $K(s)$ was an arbitrary disjunct of the tautology $\bigvee_i K(t_i^p)$, we have a Q_{k+1} -proof of

$$\bigwedge_j Sp(c_j) \rightarrow \Phi.$$

(The reader should check the fact that the construction is a m.e.b. function of the original Herbrand proof.) It remains to eliminate of the special constants.

3.36. Assume that the constants c_j are ordered according to decreasing complexity, i.e. c_1 does not occur in axioms $Sp(c_2), \dots$ etc. Let the special axiom for c_1 be $\varphi(c_1) \rightarrow (\forall x)\varphi(x)$ and replace c_1 in the formula

$$Sp(c_1) \rightarrow (Sp(c_2) \rightarrow \dots \rightarrow \Phi)$$

by a new variable y ; we get a Q_{k+2} -proof of each of the following:

$$\begin{aligned} &(\varphi(y) \rightarrow (\forall x)\varphi(x)) \rightarrow (Sp(c_2) \rightarrow \dots \Phi), \\ &(\exists y)(\varphi(y) \rightarrow (\forall x)\varphi(x)) \rightarrow (Sp(c_2) \rightarrow \dots \Phi), \\ &((\forall y)\varphi(y) \rightarrow (\forall x)\varphi(x)) \rightarrow (Sp(c_2) \rightarrow \dots \Phi), \\ &(Sp(c_2) \rightarrow \dots \Phi); \end{aligned}$$

the last formula is Q_{k+1} . By iterating this we get a Q_{k+2} -proof of Φ , q.e.d.

3.37 Remark. Note that our main trick (not necessary in Shoenfield's original proof) was to keep the complexity of the formulas involved low; $PNF(\Phi)$ need not be Q_k , but we avoid reaching $PNF(\Phi)$ by carefully inserting each quantifier in its place immediately after it has been introduced. This makes the quantifier bounded as soon as possible.

3.38 Remark. Now our proof of 3.20 has been completed (cf. 3.23). Let us indicate how to modify the whole proof in order to get a proof of Theorem

3.21. The problem is that our fixed interpretation of $I\Sigma_1$ in T may be *relative*, i.e. there may be object being non-numbers. But thanks to sequentiality, we can code sequences of arbitrary objects.

We shall sketch the general proof, the reader may elaborate details.

(1) First assume that the language of T is rich enough and contains a unary predicate $N(x)$ ranging over number (in the sense of the fixed interpretation of $I\Sigma_1$ in T), a constant $\bar{0}$, function symbols $S, +, *$ and a predicate $\leq$ having the obvious meaning in the sense of the interpretation. (If this is not the case, replace T by a conservative extension T' using the respective definitions; show that each T -cut I has a shortening J such that

$$T \vdash HCon^i(T') \rightarrow HCon^J(T).$$

(2) Define Q_o -formulas as boolean combinations of arithmetical Σ_o -formulas and arbitrary atomic formulas; then define Q_{k+1} formulas from Q_k formulas as above.

(3) We have two notions of sequences: the arithmetical notion for sequences of numbers and the notions of sequences of arbitrary objects given by the fact that T is sequential. Observe that there is a T -cut I_o such that, roughly, T proves that for lengths from I_o , the two notions of sequences of numbers coincide; in more detail, let $Seq(x)$, $(x)_y$ have its usual meaning and let $SEQ(z)$, $\beta(u, v, z)$ be as in 1.12.

Put

$$\begin{aligned} I_o(v) \equiv & N(v) \& [((\forall x)Seq(x) \rightarrow (\exists z)(SEQ(z) \\ & \& (\forall w < v)(\forall u)((x)_w = u \equiv \beta(u, w, z)) \\ & \& (\forall z)(SEQ(z) \& (\forall v < w)(\forall u)(\beta(u, w, z) \rightarrow N(u)) \\ & \rightarrow (\exists x)(Seq(x) \& (\forall w < v)(\forall u)((x)_w = u \equiv \beta(u, w, z)))]]. \end{aligned}$$

Then I_o is the desired cut.

(4) We can shorten I_o to get a cut I_1 such that, for T -terms* from I_1 , their value is uniquely determined.

Define evaluation* of variables of a term* as a particular *SEQUENCE* of objects (somehow assigning meaning to the variables* of the term*) and define a corresponding evaluating sequence for the derivation of our term*; I_1 is the collection of all $u \in I_o$ such that, for each term* $t \leq u$ and each evaluation* of its variables, each derivation* of t has a unique evaluating sequence. I_1 is a cut.

(5) By a possible further shortening we get a cut I_2 such that T proves: for terms* of the language of arithmetic and their evaluations by numbers, the new notion of the value of a term coincides with the old one (cf. I.1.64).

(6) We define satisfaction for atomic formulas $z \in I_2$ of the whole language and for Σ_o^* -formulas* $z \in I_2$ of the language of arithmetic in the obvious manner. There is a shortening J_o of I_2 such that T proves that for Q_o -formulas $z \in J_o$, satisfaction is uniquely determined and satisfies Tarski's

truth conditions. This can be iterated: for each k , there is a cut J_k and a satisfaction for Q_k -formulas^{*} such that T proves its usual properties.

(7) Given k and Q_k , J_k has a shortening K_k such that T proves that Q_k -proofs from true formulas are truth-preserving. Then we can continue as in the proof 3.20: there is a m.e.b. function transforming each H -proof^{*} of a Q_k -formula^{*} into a Q_{k+2} -proof of that formula. This completes our proof sketch of 3.21.

(d) Cuts and Interpretations

This subsection contains a proof of the following theorem.

3.39 Theorem. Let S, T be theories, finitely axiomatized and $T \supseteq I\Sigma_1$ and sequential. Then S is interpretable in T iff there is a T -cut I such that $T \vdash HCon^I(S^*)$.

Proof. First let us discuss the implication $\Rightarrow$. The proof is a generalization of the proof of 3.20, thus a modification of our proof of 2.21. We describe the necessary changes. First, simplify the notion of Q_k -formulas: since we do not assume that the language of S contains the language of arithmetic, let Q_0 -formulas be all atomic S -formulas; Q_{k+1} -formulas are boolean combinations of $\exists Q_k$ -formulas. Thus Q_k -formulas are particular S -formulas. Show by induction on k , using the given interpretation, that in T there is a cut I_k and a satisfaction $Sat_{Q,k}$ for Q_k -formulas such that T proves Tarski's truth conditions and thus T proves the following "snowing"-snowing-lemma: for each S -formula $\varphi(x, \dots)$ which is a Q_k -formula,

$$T \vdash (\chi(x) \& \dots) \rightarrow (\varphi^*(x, \dots) \equiv Sat_{Q,k}(\varphi, [x, \dots])).$$

(Here $*$ is the interpretation).

There is a shortening J_k of I_k such that $T \vdash (Q_k$ -proofs $z \in I$ preserve $Sat_{Q,k}$ -truth) and there is a shortening K_k of J_k such that T proves that each H -proof^{*} (from no special axioms) $u \in K_{k+2}$ of a Q_k -formula φ determines a Q_{k+2} -proof $z \in J_{k+2}$ of φ . Consequently T proves the following: if (the concatenation of all elements of) S^* is Q_k then K_{k+2} contains no H -proof^{*} of $\neg S^*$ (recall that $T \vdash (S^* \text{ is true})$ since $T \vdash S^*$). Thus the proof of $\Rightarrow$ is complete. The rest of the subsection contains a proof of $\Leftarrow$. $\square$

3.40 Lemma. Let $T \supseteq I\Sigma_1$. For each T -cut I there is a T -cut J such that $T \vdash I \subseteq J$ and $T \vdash (J \text{ is closed under concatenation of sequences})$.

Proof. Let us work in T . Let $s, t \leq x$; then $lh(s), lh(t) \leq |x|$ and, for each i , $(s)_i, (t)_i \leq |x|$. If w is the concatenation of s, t , i.e. $w = s \frown t$, then

$lh(u) \leq 2|x|$ and each member is $\leq |x|$, thus, by 3.13, $w \leq 2^{(3(x)+1)^2}$. An elementary computation shows that

$$\omega_2(x) = 2^{1/4 * (|x|-1)^2};$$

thus it suffices to assume that J is closed under ω_2 (cf. 3.5). $\square$

3.41 Corollary. Let L be a finite language, let $T \supseteq I\Sigma_1$, let I be a T -cut. Then there is a T -cut J such that $T \vdash J \subseteq I$ and T proves that for each function symbol F and predicate symbol P of L^* , of arity n , whenever $t_1, \dots, t_n$ are in J then so is $F(t_1, \dots, t_n)$ and $P(t_1, \dots, t_n)$.

3.42 Proof of 3.39 $\Leftarrow$. Our starting situation is: S is a finite theory (assume S is a one-element set), $T \supseteq I\Sigma_1$ is sequential, I is a T -cut such that $T \vdash HCon^*(S^*)$. Thus $T \vdash$ (in I , there is no propositional *proof*^{*} of any disjunction of instances of $\neg Sk(S^*)$), cf. the definition I.4.11 of $He^*(\neg S^*)$.

Define, in T , a function F described as follows: for each x , first take the conjunction κ_x of all closed instances of $Sk(S^*)$ less than x , then construct the truth table for κ_x . If there is an evaluation of atoms making κ_x propositionally true then $F(x)$ is such an evaluation (e.g. the least one). Otherwise $F(x)$ is a propositional proof of $\neg \kappa_x$. Check that F is Δ_1 and m.e.b. in T . Let J be a T -cut shorter than I and such that $T \vdash x \in J \rightarrow F(x) \in I$. Then T proves that, for each $x \in J$, the conjunction κ_x (of all closed instances of $Sk^*(S^*)$ less than x) is propositionally satisfiable.

(2) What follows resembles König's lemma; but since we do not have enough induction we have to replace infinity in the sense of the universe by infinity in the sense of a cut. Namely, let us work in T , let L_1 be the language of $Sk(S)$ and let $V(e)$ mean that, for some $x \in J$, e is an evaluation of closed L_1^* -atoms less than x , (say, an x -evaluation) making κ_x propositionally true. V is, in the obvious sense, a dyadic tree and is, so to speak, J -infinite: for each $x \in J$, some x -evaluation is in V . An evaluation $e \in V$ for κ_x is said to have J -unboundedly many prolongations in V if, for each $y \in J$, $y > x$, there is an evaluation $e' \in V$ for κ_y such that $e \subseteq e'$. Clearly, if $e \in V$ is an x -evaluation and has J -unboundedly many prolongations in V then there is a least $(x+1)$ -evaluation $e' \in V$ extending e and having J -unboundedly many prolongations in V (since there are at most two $(x+1)$ -evaluations prolonging e). Write, for a moment, $LP(e)$ for e' . An x -evaluation $e \in V$ is leftmost (i.e. leftmost having J -unboundedly many prolongations) if, for each $y < x$, $e \upharpoonright (y+1) = LP(e \upharpoonright y)$. Clearly, if an x -evaluation $e \in V$ is leftmost $(x+1)$ -evaluations prolonging e .

Put $J_1(x) \equiv J(x)$ & there is a unique leftmost x -evaluation $e \in V$.

Clearly, J_1 is a T -cut and $T \vdash J_1 \subseteq J$; we may define a function B on J_1 assigning to each closed L_1 -atom in J_1 its truth (0 or 1) in accordance with the unique leftmost x -evaluation (for any x satisfactorily large).

(3) Shorten J_1 to J_2 such that $T \vdash (J_2 \text{ is closed under concatenation})$ and restrict B to J_2 . Thus, T -provably, if $t_1, \dots, t_n$ are closed terms in J_2 and $\bar{F}$, is a function symbol from L_1 (of arity n) then the term $\bar{F}(t_1, \dots, t_n)$ is also in J_2 . Similarly for a predicate.

(4) We put

$$\chi(x) \equiv x \text{ is a closed } L_1^\bullet\text{-term in } J_2;$$

for each function symbol F of L , let

$$\psi_F(x_1, \dots, x_n, y) \equiv y \text{ is the term } \bar{F}(x_1, \dots, x_n),$$

(application $^\bullet$ of $\bar{F}$ to $x_1, \dots, x_n$) and for each predicate P of L , let

$$\psi_P(x_1, \dots, x_n) \equiv B \text{ assigns the value 1 to the closed atom } \bar{P}(x_1, \dots, x_n).$$

This defines our interpretation; it suffices to verify that it is an interpretation of $Sk(S)$ in T . But this is now clear: each ψ_F defines a total n -ary function on the set of all closed terms (which is non-empty). Verify (by induction outside T) that, for each L_1 -term $t(x_0, \dots, x_n)$, T proves

$$\chi(u_0) \& \dots \& \chi(u_n) \rightarrow [t^*(u_0, \dots, u_n) = \bar{t}(n_0/x_0, \dots, u_n/x_n)]$$

(cf. I.1.67), and, for each open L_1 -formula $\varphi(x_0, \dots, x_n)$, T proves

$$\chi(u_0) \& \dots \& \chi(u_n) \rightarrow [\varphi^*(u_0, \dots, u_n) \equiv \text{with respect to } B, \\ \varphi(u_0/x_0, \dots, u_n/x_n) \text{ has truth value 1}].$$

(Note that you always need a piece of B which is a finite set.) Since T proves that each closed instance of $Sk(S^\bullet)$ has truth value 1 under B , we get $T \vdash [(\forall \dots) Sk(S)]^\bullet$ this completes the proof of 3.39. $\square$

3.43 Remark. If S contains an equality predicate we may be interested in an interpretation with absolute equality, i.e. such that $T \vdash \chi(x) \& \chi(y) \rightarrow (x = y \equiv x =^* y)$. The above construction is easily refined to get an interpretation with absolute equality: in T , call a closed term u B -minimal if for each $u' < u$ such that $\chi(u')$, B gives value 0 to the formula $^\bullet$ $u =^* u'$. Note that with each u such that $\chi(u)$ we may associate an $u' = MT(u)$ such that u' is minimal and $B(u =^* u') = 1$. (Again, this because to decide $u =^* u'$ for all $u \leq u'$ we need to know only a set-piece of $B^\bullet$). Thus replace $\chi(u)$ by $(\chi(u) \& u \text{ is } B\text{-minimal})$ and make appropriate changes – details left to the reader.

3.44 Remark. For PA , we get the following as a corollary: A finite theory S is interpretable in PA iff $PA \vdash Con^\bullet(S^\bullet)$. Clearly, this could be obtained directly from other previous results.

However, we can now generalize investigations of Sect. 2 (e), concerning pure extensions of PA , i.e. extensions of PA in the language of PA to arbitrary sequential theories T containing PA (i.e. PA is interpreted in T , the interpretation being fixed) and having full induction, i.e. if $N(x)$ in the predicate “ x is a number” and $\varphi(x)$ is any formula of the language of T then T proves

$$\varphi(\bar{0}) \& (\forall x)(N(x) \& \varphi(x) \rightarrow \varphi(x+1)) \rightarrow (\forall x)(N(x) \rightarrow \varphi(x)).$$

Note that if T is sequential and has full induction (with respect to the given interpretation of Q in N) then T contains PA .

3.45 Theorem. Let $T \in \Delta_1$ be a sequential theory having full induction, let $S \in \Delta_1$ be another theory. Then the following are equivalent:

- (i) S is interpretable in T .
- (ii) S is locally interpretable in T .
- (iii) For each k , $T \vdash \text{Con}^*((S \upharpoonright k)^*)$.
- (iv) There is a binumeration σ of S such that $T \vdash \text{Con}^*(\sigma)$.

3.46 Theorem. If T and S are as above and S is sequential and has full induction then S is interpretable in T iff S is Π_1 -conservative for T .

To prove 3.46, just check proofs of 2.39 and 2.40. You have to use the following two facts:

3.47 Lemma. If T is sequential and has full induction then T is reflexive.

Proof. By 3.21, for each finite $T_0 \supseteq T$ we have $T \vdash \text{HCon}^*(T_0^*)$ (since there are no proper T -cuts); but T contains $I\Sigma_1$ and therefore $T \vdash \text{Con}^*(T_0^*) \equiv \text{HCon}^*(T_0^*)$. $\square$

3.48 Remark. In the proof of (iv) $\Rightarrow$ (i) observe that we get an interpretation of S in T such that all S -objects are interpreted as some numbers in T , thus we may apply the least number principle and get an interpretation absolute with respect to equality.

4. Partial Conservativity and Interpretability

4.1. In the present section we shall investigate the notion of Γ -conservativity (Γ being a class of formulas) and interpretability as means of comparing theses. Both notions were defined in the present chapter, Sect. 1 (a); and for pure extensions of PA (more generally, for sequential theories with full induction) we already have a result saying that S is interpretable in T iff

S is Π_1 -conservative for T (see 2.40 and 3.46). Now we want to remove the assumption of full induction and, in particular, get results for finitely axiomatized theories. Recall that we have the characterization 3.39 of interpretability of a finite S in a sequential $T \supseteq I\Sigma_1$ ($T \vdash HCon^{\bullet I}(S^{\bullet})$ for some T -cut I ; this will be used repeatedly.) We shall see that results concerning partial conservativity do not depend on any assumption of finite axiomatization, but properties of interpretability in finite theories differ drastically from properties of interpretability in theories with full induction.

We shall particularly focus our attention on pairs S, T of theories (containing $I\Sigma_1$, say) such that S results from T by adding one axiom, thus S is $(T + \varphi)$ for some φ . If φ is independent of T , i.e. neither provable nor refutable in T , it is natural to ask whether S is interpretable in T and how conservative S is over T . Instead of saying that $(T + \varphi)$ is Γ -conservative over T we say that φ is Γ -conservative over T ; similarly for interpretability. In subsection (a) we shall ask these questions for some prominent formulas (Gödel's and Rosser's formula); subsection (b) contains some general theorems on partial conservativity and in subsection (c) these theorems are applied and related to interpretability. All this is a possible answer to the question what more we can say concerning axiomatic systems of arithmetic than that they are all incomplete.

(a) Some Prominent Examples

4.2. In the whole subsection, T denotes a theory containing $I\Sigma_1$ and τ is a formula Δ_1 in $I\Sigma_1$ defining T (thus τ binumerates T in T). We shall often write $Con_T^{\bullet}$ instead of $Con_{\tau}^{\bullet}$. $Con_T^{\bullet}$ is called Gödel's formula; recall that, by 2.22, $Con_T^{\bullet}$ is equivalent to the self-referential formula ν such that $T \vdash \nu \equiv \neg Pr_T(\bar{\nu})$, i.e. to Gödel's fixed point. Clearly, $Con_T^{\bullet}$ is a Π_1 -sentence. Rosser's formula ρ is the self-referential formula such that (cf. 2.9).

$$T \vdash \rho \equiv (\exists y)(Proof_T^{\bullet}(y, \neg \rho) \& (\forall z \leq y) \neg Proof_T^{\bullet}(z, \bar{\rho})),$$

i.e. ρ says "there is a proof y of my negation such that there is no proof of me beneath y ". Thus ρ is a Σ_1 -sentence. These formulas played a prominent role in Gödel's incompleteness theorems (cf. 2.8, 2.10, 2.22); now we shall discuss their properties concerning partial conservativity and interpretability.

4.3. Recall the notion of H -provability and H -proofs (H for Herbrand) in a finite theory T and the sentence $HCon^{\bullet}(T)$ expressing consistency with respect to Herbrand proofs.

4.4 **Definition.** Let T be a finite theory. An H -Rosser formula for T is a self-referential formula ρ such that

$$T \vdash \rho \equiv (\exists y)(HProof_T^{\bullet}(y, \neg \rho) \& (\forall z \leq y)(\neg HProof_T^{\bullet}(z, \bar{\rho})).$$

This is like the Rosser formula but with Herbrand proofs instead of (Hilbert) proofs.

Our results are summarized in the following

4.5 Theorem. Let $T \in \Delta_1$ be a consistent theory, $T \supseteq I\Sigma_1$.

- (1) Gödel's formula $Con_T^\bullet$ is not interpretable in T ; its negation $\neg Con_T^\bullet$ is interpretable in T .
- (2) $\neg Con_T^\bullet$ is Π_1 -conservative over T ; $Con_T^\bullet$ is Σ_1 -conservative over T iff T is Σ_1 -sound (i.e. each provable Σ_1 -sentence is true).
- (3) Rosser's formula is Π_1 -conservative; $\neg\rho$ is Σ_1 -conservative iff T is Σ_1 -sound. (The same for H -Rosser's formula assuming T finite.)
- (4) If T is sequential and has induction for all formulas then neither the Rosser's formula ρ nor $\neg\rho$ is interpretable in T .
- (5) But if T is sequential and finite and $H\rho$ is H -Rosser's formula then both $H\rho$ and $\neg H\rho$ are interpretable in T .

4.6 Remark. (1) There is an open problem if for T as in (5) the usual Rosser's formula ρ and/or its negation is interpretable in T .

(2) The rest of the subsection contains a proof of 4.5.

*

4.7 Lemma. Let $S, T \in \Delta_1$, $S, T \supseteq I\Sigma_1$, let S have a finite language and let S be interpretable in T . Then, for each Σ_1 definition τ of T there is a Σ_1 definition σ of S such that

$$T \vdash Con_\tau^\bullet \rightarrow Con_\sigma^\bullet.$$

Proof. Let i be the mapping of S -formulas into T -formula induced by the interpretation; i is Δ_1 . Copy the definition in T : we get a function $i^\bullet$, Δ_1 in T , such that, for each $\psi \in S$, T proves $i(\psi)$ and hence $T \vdash Pr_\tau^\bullet(i^\bullet(\bar{\psi}))$. Let σ be a Σ_1 definition of S . We cannot claim

$$T \vdash (\forall x)(\sigma_o(x) \rightarrow Pr_T^\bullet(i^\bullet(x)));$$

but take $\sigma(x)$ to be $\sigma_o(x) \ \& \ Pr_T^\bullet(i^\bullet(x))$. It is easy to check that σ is the desired definition of S . (Note that if τ is a Σ_1 binumeration of T in T then σ is a Σ_1 binumeration of S in T .) $\square$

4.8 Lemma. Let T be as above and consistent; then the formula $Con_T^\bullet$ is not interpretable in T .

Proof. Assume that the theory $(T \vdash Con_T^\bullet)$ is interpretable in T ; then, by the preceding lemma, there is a Σ_1 -definition σ of $T + Con_T^\bullet$ such that

$T \vdash \text{Con}_T^* \rightarrow \text{Con}_\sigma^*$, i.e. $(T + \text{Con}_T^*) \vdash \text{Con}_\sigma^*$, which contradicts Gödel's second incompleteness theorem (see 2.22). $\square$

4.9 Lemma. Let T be as above; then $\neg \text{Con}_T^*$ is interpretable in T and is Π_1 -conservative over T .

Proof. We show that $(T + \neg \text{Con}_T^*)$ is interpretable in, and Π_1 conservative for, $(T + \text{Con}_T^*)$. This gives interpretability in T by 1.8 and conservativity over T using the rule of proof by cases.

By Gödel's second incompleteness theorem, 2.22, copied inside T , we get $T \vdash \text{Con}_T^* \rightarrow \neg \text{Pr}_T(\overline{\text{Con}_T^*})$, i.e.

$$T \vdash \text{Con}^*(T) \rightarrow \text{Con}^*(T + \overline{\neg \text{Con}_T^*})$$

(or, more pedantically, if τ is a Σ_1 definition of T and $\tau'(x)$ is $\tau(x) \vee x = \overline{\neg \text{Con}_T^*}$ then $T \vdash \text{Con}_\tau^* \rightarrow \text{Con}_{\tau'}^*$). Thus take $(T + \text{Con}_T^*)$. By the low arithmetized completeness theorem, in $(T + \text{Con}_T^*)$ we may define a full low model M of $T + \overline{\neg \text{Con}_T^*}$; seen from outside, we just get an interpretation of $T + \neg \text{Con}_T^*$ in $T + \text{Con}_T^*$. This proves interpretability. Moreover, the interpretation is very well behaved, since everything is low $\Sigma_0^*(\Sigma_1)$; thus we may imitate the proof in 2.44 and construct inside $T + \text{Con}_T^*$ a $\Sigma_0^*(\Sigma_1)$ -embedding F of the universe onto an initial segment of M . (Induction for $\Sigma_0^*(\Sigma_1)$ used!) Then conclude as in 2.44. $\square$

4.10 Lemma. If $T \supseteq I\Sigma_1$ is Σ_1 -sound then each non-refutable Π_1 -sentence is Σ_1 -conservative over T .

Proof. Let $\pi \in \Pi_1$, $\sigma \in \Sigma_1$, $\neg \pi$ unprovable in T . If $(T + \pi) \vdash \sigma$ then $T \vdash \neg \pi \vee \sigma$ and $\neg \pi \vee \sigma$ is Σ_1 in T . From the non-refutability of π we get $N \models \pi$ by Σ_1 -completeness; on the other hand, from $T \vdash \neg \pi \vee \sigma$ we get $N \models \neg \pi \vee \sigma$ by Σ_1 -soundness. This gives $N \models \sigma$ and $T \vdash \sigma$. $\square$

4.11 Corollary. If $T \in \Delta_1$ contains $I\Sigma_1$ and is Σ_1 -sound then Con_T^* is Σ_1 -conservative.

We shall prove the converse implication later. Now we turn to Rosser's sentences.

4.12 Lemma. Let $T \in \Delta_1$, $T \supseteq I\Sigma_1$, T consistent. Then Rosser's sentence ρ_T is Π_1 -nonconservative.

Proof. ρ evidently implies the following Π_1 -sentence

$$(\forall z)(\text{Proof}^*(\bar{\rho}, z) \rightarrow (\exists y < z)\text{Proof}^*(\neg \bar{\rho}, z)).$$

Denote it by π . If $T \vdash \pi$ then $T \vdash Pr_T^*(\bar{\rho}) \rightarrow \rho$, since, by Löb's theorem 2.25, $T \vdash \rho$, which is impossible. Thus $(T + \rho)$ proves π but T does not. $\square$

4.13 Remark. The same holds for H -Rosser's formula (provided T is finite).

4.14 Corollary. If T is a consistent sequential theory with full induction then neither ρ nor $\neg\rho$ are interpretable (i.e. T interprets neither $(T + \rho)$ nor $T + \neg\rho$).

This follows by 3.46, 4.12 and the fact that $\neg\rho$ is an unprovable Π_1 formula. Similarly for H -Rosser. On the other hand, we have the following:

4.15 Lemma. If $T \supseteq I\Sigma_1$ is consistent, finitely axiomatized and sequential then both the H -Rosser sentence and its negation are interpretable in T .

Proof. For simplicity assume T to be just a sentence. Let ρ be the Rosser sentence. We use Theorem 3.39. It suffices to interpret $(T + \rho)$ in $(T + \neg\rho)$ and conversely. We show that $(T + \rho)$ is interpretable in $(T + \neg\rho)$, the existence of the converse interpretation is proved analogously.

Let J be a $(T + \neg\rho)$ -cut such that $(T + \neg\rho) \vdash HCon^{*J}(T + \neg\rho)$. Let us work in $(T + \neg\rho)$. In J there is no H -proof * of $\neg(T + \neg\rho)$; and $\neg\rho$ says that for each H -proof * of $\neg\bar{\rho}$ in T (i.e. for each H -proof * $\neg(T \& \neg\rho)$) there is a smaller H -proof * of $\bar{\rho}$ in T (i.e. a H -proof * of $\neg(T \& \neg\rho)$). Consequently, J does not contain any H -proof * of $\neg(T \& \neg\rho)$ and we get $HCon(T + \neg\rho)$.

The provability of the last formula in $(T + \neg\rho)$ gives an interpretation of $(T + \neg\rho)$ in $(T + \rho)$, hence an interpretation of $(T + \rho)$ in $(T + \neg\rho)$ by 3.39. $\square$

4.16 Corollary and Remark. If $T \supseteq I\Sigma_1$ is consistent, finite and sequential then there is a Π_1 formula φ such that both φ and $\neg\varphi$ are interpretable in T (once more, this means that the theory $(T + \varphi)$ is interpretable in T and so is $(T + \neg\varphi)$).

Note that this is not true for sequential theories with full induction (like PA and ZF) since in such a theory each interpretable Π_1 -sentence is provable (cf. 2.40). But one can find a $\varphi \in \Delta_2$ such that both φ and $\neg\varphi$ is interpretable.

On the other hand, take ACA_0 for T ; assumptions of our corollary are satisfied. The corresponding formula φ has the following properties: $(ACA_0 + \varphi)$ is interpretable in ACA_0 but $(PA + \varphi)$ is not interpretable in PA (since φ is unprovable). Similarly for Gödel-Bernays and Zermelo-Fraenkel set theories GB and ZF . In the next section we shall obtain a Σ_1 sentence φ with converse properties: $(PA + \varphi)$ is interpretable in PA but $(ACA_0 + \varphi)$ is not in ACA_0 .

4.17 Lemma. Let $T \supseteq I\Sigma_1$ be consistent.

- (1) Con_T^* is Σ_1 -conservative over T iff T is Σ_1 -sound.
- (2) Let ρ be the Rosser (or H -Rosser) Σ_1 -sentence. Then $\neg\rho$ is Σ_1 -conservative iff T is Σ_1 -sound.

Proof. We shall prove (1). The proof will be a preparation for generalized Rosser sentences studied in the next subsection; there we also prove assertion (2).

The implication $\Leftarrow$ is evident: if $\sigma \in \Sigma_1$, T is Σ_1 -sound and $T \vdash \text{Con}_T^\bullet \rightarrow \rho$ then $N \models \text{Con}_T^\bullet \rightarrow \rho$, $N \models \rho$ and $T \vdash \rho$.

Now assume T not to be Σ_1 -sound and let $\varphi(y)$ be a Σ_0 -formula such that $N \models \neg(\exists y)\varphi(y)$ but $T \vdash (\exists y)\varphi(y)$. Take the following self-referential formula:

$$T \vdash \xi \equiv (\exists y)(\text{Proof}^\bullet(\neg\xi, y) \vee \varphi(y) \& (\forall z \leq y)\neg\text{Proof}^\bullet(\bar{\xi}, z)).$$

Note that $T \vdash \neg\varphi(\bar{k})$ for each k .

(i) We show that T does not prove ξ . Assume $T \vdash \xi$; we show that T is inconsistent. Let d be a proof of ξ in T ; then $T \vdash \text{Proof}^\bullet(\bar{\xi}, \bar{d})$ and for each $k < d$, $T \vdash \neg\varphi(\bar{k})$.

From $T \vdash \xi$ we get

$$\begin{aligned} T \vdash (\exists y < \bar{d})\text{Proof}^\bullet(\neg\xi, y), \\ T \vdash \bigvee_{k < d} \text{Proof}^\bullet(\neg\xi, \bar{k}) \end{aligned}$$

and since each formula $\text{Proof}^\bullet(\neg\xi, \bar{k})$ is decidable in T there is a k such that

$$T \vdash \text{Proof}^\bullet(\neg\xi, \bar{k}).$$

Thus k is a proof of $\neg\xi$ in T , thus $T \vdash \neg\xi$ and T is inconsistent. This shows that ξ is unprovable in T provided T is consistent.

(ii) We now prove $T + \text{Con}_T^\bullet \vdash \xi$ or, equivalently, $T + \neg\xi \vdash \neg\text{Con}_T^\bullet$. Let us work in $(T + \neg\xi)$. Since $(\exists y)\varphi(y)$, let y_0 be the least y such that $\varphi(y)$. From $\neg\xi$ we get $(\exists z \leq y_0)\text{Proof}^\bullet(\bar{\xi}, z)$; let $\text{Proof}^\bullet(\bar{\xi}, z_0)$. Then $\text{Proof}^\bullet(\bar{\xi}, z_0) \& (\forall y < z_0)\neg\varphi(y)$, hence inside T we are in the same situation as we were in (i) outside. Copy the reasoning from (i) into T ; we get

$$\text{Pr}^\bullet(\overline{\text{Proof}^\bullet(\bar{\xi}, z_0)}) \& (\forall y < z_0)\text{Pr}^\bullet(\neg\varphi(y)).$$

From $\text{Pr}^\bullet(\xi)$ we get

$$\text{Pr}^\bullet\left(\bigvee_{y < z_0} \overline{\text{Proof}^\bullet(\neg\xi, y)}\right),$$

hence

$$\begin{aligned} (\exists y < z_0)\text{Pr}^\bullet(\overline{\text{Proof}^\bullet(\neg\xi, y)}), \\ (\exists y < z_0)\text{Pr}^\bullet(\neg\xi, y), \end{aligned}$$

thus we get $\text{Pr}^\bullet(\bar{\xi})$ and $\text{Pr}^\bullet(\neg\bar{\xi})$, thus $\neg\text{Con}^\bullet$. □

We have presented a rather detailed proof; similar proofs below will mostly be less detailed.

(b) General Theorems on Partial Conservativity; Some Fixed-Point Theorems

In this subsection we shall state and prove some basic facts on partially conservative sentences. Our proofs will use self-reference as a very basic means. Some methods of construction of self-referential sentences will be isolated and properties of constructed formulas will be stated in general theorems called usually fixed-point theorems. It is remarkable that most of our constructions are some generalizations of Rosser's formula, i.e. are based on witness comparisons. We first introduce some notations and formulate a simple technical lemma, then we formulate our basic results (inclusive the fixed-point theorems) and finally we elaborate proofs. In the whole subsection, T is a consistent theory, $T \in \Delta_1$, $T \supseteq I\Sigma_1$.

4.18 Definition Let $\alpha(u)$, $\beta(u)$ be T -formulas, let Δ be $(\exists u)\alpha(u)$ and let ∇ be $(\exists u)\beta(u)$. In T , call each a satisfying $\alpha(u)$ a witness for Δ and similarly for ∇ (cf. 2.9). By $\Delta \prec \nabla$ we denote the witness comparison formula

$$(\exists u)(\alpha(u) \& (\forall v \leq u) \neg \beta(v))$$

(there is a witness for Δ less than each witness for ∇); similarly for $\Delta \preceq \nabla$ (replace $(\forall v \leq u)$ by $(\forall v < u)$). The formula Δ is called the *antecedent* of $\Delta \prec \nabla$ and Δ its *succedent*. Similarly for $\Delta \preceq \nabla$. Note that Δ and ∇ may contain free variables as parameters. (Thus, for example, Rosser's formula is a formula ρ such that

$$T \vdash \rho \equiv Pr^\bullet(\neg \rho) \prec Pr^\bullet(\rho).)$$

4.19 Remark. Rosser's formula is Σ_1 in T , since we assume $T \supseteq I\Sigma_1$. Recall that in $B\Sigma_n$, Σ_n formulas are closed under bounded universal quantifiers. We shall deal with formulas comparing witnesses of Σ_n -formulas but continue to assume only $I\Sigma_1$. Therefore we shall have to pay some attention to the arithmetical complexity of the resulting formulas. The following simple lemma will be extremely useful.

4.20 Lemma. Let $\varphi(x)$ be a Σ_n -formula. There is a Σ_n -formula $\psi(y)$ such that

- (1) $(\forall k)I\Sigma_1 \vdash \psi(\bar{k}) \equiv (\forall x \leq \bar{k})\varphi(x),$
- (2) $I\Sigma_1 \vdash \psi(y) \rightarrow (\forall x \leq y)\varphi(x).$

Proof. For $n = 0$ the assertion is trivial. For $n \geq 1$ and $\varphi(x)$ of the form $(\exists u)\alpha(x, u)$, where α is Π_{n-1} let $\psi(y)$ be

$$(\exists s)(Seq(s) \& (\forall x \leq y)\alpha(x, (s)_x)). \quad \square$$

4.21 Definition. (1) If $\varphi(x)$ and $\psi(y)$ are as above then we denote $\psi(y)$ by $[(\forall x \leq y)\varphi(x)]^{\#,\Sigma,n}$ or $[(\forall x \leq y)\varphi(x)]^{\#}$ (if Σ and n is clear from the context). Dually we define $[(\exists x \leq y)\varphi(x)]^{\#,\Pi,n}$ for φ being Π_n (or, briefly, $[(\exists x \leq y)\varphi(x)]^{\#}$).

(2) Let α and β be as by 4.18 and assume that β is Π_n . Let β' be the natural Σ_n -form of $\neg\beta$. Then $\Delta \prec^{\#} \nabla$ is the formula

$$(\exists u)(\alpha(u) \& [(\forall v \leq u)\beta'(v)]^{\#,\Sigma,n});$$

the notation $\Delta \prec^{\#} \nabla$ will be used only in situations where n is clear from the context. Similarly for $\Delta \preceq^{\#} \nabla$; the words “antecedent” and “succedent” have the obvious meaning.

4.22 Remark. We shall deal with self-referential formulas ξ such that

$$T \vdash \xi \equiv \Delta(\overline{\neg\xi}) \prec^{\#} \nabla(\overline{\xi});$$

more generally, there may be another parameter:

$$T \vdash \xi \equiv \Delta(\overline{\neg\xi}, \overline{k}) \prec^{\#} \nabla(\overline{\xi}, \overline{k})$$

(k natural number). Such formulas will be extremely useful; almost always the antecedent and/or the succedent will contain some proof predicate. Let us mention $Pr_T^{\bullet}$ and $HPr_T^{\bullet}$ (Herbrand provability, assuming T finite.) Our general theorems will be obtained by combining self-reference and witness comparison with definable cuts and partial truth definitions. Recall that satisfaction for $\Sigma_n^{\bullet}$ -formulas^{*} and $\Pi_n^{\bullet}$ -formulas ($Sat_{\Sigma,n}$, $Sat_{\Pi,n}$) were defined in $I\Sigma_1$ already in I.1.75–76; in Chap. I, Sect. 2 we used them to prove finite axiomatizability of $I\Sigma_n$ and $B\Sigma_{n+1}$ and studied satisfaction for relativized hierarchies. In Chap. I, Sect. 4 we showed that these satisfaction relations are just a particular case of the general notion of partial satisfaction, namely, satisfaction for the standard model (defined inside $I\Sigma_1$). In the present chapter we already met Σ_1 -satisfaction in Sect. 3 (Theorem 3.21 saying that for suitable (finite) theories T , there is a T -cut J such that T proves $(\forall u \in Tr(\Sigma_1))HCon^{\bullet J}(T + u)$ (for any true $\Sigma_1^{\bullet}$ -sentence, in J there is no Herbrand-proof^{*} of $\neg(\wedge T \& u)$). We shall now make systematic use of provability (H -provability) in extensions of T resulting by adding a true Γ -formula^{*} where Γ is a class of formulas^{*}, say $\Sigma_n^{\bullet}$ or $\Pi_n^{\bullet}$. This leads us to the following.

4.23 Definition and Convention. Let $n \in N$, let Γ be Σ_n or Π_n . We make the following definition in IS_1 : $Pr_{T,\Gamma}^\bullet(x)$ (x is provable in T from a true Γ -sentence) if

$$(\exists y \in \Gamma)(Tr_\Gamma(y) \& Pr_{(T+y)}^\bullet(x)).$$

If T is finite we define $HPr_{T,\Gamma}^\bullet(x)$ in the obvious way.

Convention. (1) To simplify notation, in the rest of this section we drop the index T and the dot; thus we shall write $Pr(x)$, $Pr_\Gamma(x)$ instead of $Pr_T^\bullet(x)$, $Pr_{T,\Gamma}^\bullet(x)$ etc.

(2) If Δ_i is $(\exists u)\alpha_i(u)$ and ∇_i is $(\exists v)\beta_i(v)$ ($i = 1, 2$) then $(\Delta_1 \vee \nabla_2) \prec (\Delta_1 \vee \nabla_2)$ means the formula saying "there is a witness for $\alpha_1 \vee \alpha_2$ less than each witness for $\beta_1 \vee \beta_2$ ", similarly for $\prec^\#$ instead of $\prec$.

4.24 Definition. (1) φ is *hereditarily Γ -conservative* over T if, for each T_0 such that $IS_1 \subseteq T_0 \subseteq T$, φ is Γ -conservative over T_0 .

(2) φ is *doubly Γ -conservative* over T if φ is Γ -conservative over T and $\neg\varphi$ is $\tilde{\Gamma}$ -conservative over T (where $\tilde{\Gamma}$ is the dual class of Γ).

We shall now formulate three general theorems on partial conservativity. Note that we continue to assume that T is consistent, $T \in \Delta_1$ and $T \supseteq IS_1$.

4.25 Theorem. For each $n \geq 1$ there is (1) a hereditarily Π_n -conservative Σ_n -sentence, (2) a hereditarily Σ_n -conservative Π_n -sentence, (3) a doubly Π_n -conservative Σ_n -sentence (its negation is thus a doubly Σ_n conservative Π_n -sentence).

Examples (Γ is Σ_n , Λ is Π_n):

- (1) ξ such that $IS_1 \vdash \xi \equiv Pr_\Gamma(\neg\bar{\xi}) \prec^\# Pr_\Lambda(\bar{\xi})$,
- (2) $(\neg\xi)$ such that $IS_1 \vdash \xi \vdash Pr_\Gamma(\neg\bar{\xi}) \prec^\# Pr_\Lambda(\bar{\xi})$,
- (3) ξ such that $IS_1 \vdash \xi \subseteq Pr_\Gamma(\neg\bar{\xi}) \prec^\# Pr_\Lambda(\bar{\xi})$.

If T is Σ_n -sound we may take in (1) a ξ such that $IS_1 \vdash \xi \equiv Pr_\Gamma(\neg\bar{\xi})$.

4.26 Theorem (on non-separability). Let Γ be Σ_n or Π_n ($n \geq 1$), let Th be the set of all theorems of T , $Consv(\Gamma)$ and $hConsv(\Gamma)$ the set of all Γ -conservative and hereditarily Γ -conservative sentences respectively, $NRef$ the set of all the sentences non-refutable in T . Then obviously

$$Th \subseteq hConsv(\Gamma) \subseteq Consv(\Sigma_1) \cap Consv(\Pi_1) \subseteq NRef$$

and there is no set X such that

- (1) X is Δ_1 and $Th \subseteq X \subseteq NRef$, or
- (2) X is Π_1 and $Th \subseteq X \subseteq Consv(\Gamma)$, or
- (3) X is Σ_1 and $hConsv(\Gamma) \subseteq X \subseteq NRef$, or
- (4) X is Σ_2 , $\Gamma \supseteq \Sigma_1$ and $hConsv(\Gamma) \subseteq X \subseteq Consv(\Sigma_1)$, or
- (4') X is Σ_2 , $\Gamma \supseteq \Pi_1$ and $hConsv(\Gamma) \subseteq X \subseteq Consv(\Pi_1)$.

4.27 Theorem (Π_2 -completeness). For each $n \geq 1$ and $\Gamma = \Sigma_n$ or Π_n , both $\text{Consv}(\Gamma)$ and $\text{hConsv}(\Gamma)$ is Π_2 -complete.

4.28 Remark. (1) Concerning 4.26, observe that Th is Σ_1 , NRef is Π_1 and $\text{Consv}(\Gamma)$ is Π_2 . Thus the result is optimal: in 4.26 (1) Δ_1 can be replaced by neither Σ_1 nor Π_1 , in 4.26 (2) Π_1 cannot be replaced by Σ_1 , in 4.26 (3) Σ_1 cannot be replaced by Π_1 , and in 4.26 (4,4') Σ_2 cannot be replaced by Π_2 .

(2) Recall from recursion theory that a set $X \subseteq N$ is Π_2 -complete if $X \in \Pi_2$ and each Π_2 set Y is Δ_1 -reducible to X , i.e. for some Δ_1 total function F , $Y = \{n \mid F(n) \in X\}$.

(3) We shall present two general fixed point theorems that form the main means of proofs of the preceding theorems.

4.29 Sheperdson-Smoryński's Fixed Point Theorem. Let Φ, Ψ be Σ_1 formulas.

- (1) Let $I\Sigma_1 \vdash \xi \equiv [(Pr(\neg\xi) \vee \Phi) \prec (Pr(\xi) \vee \Psi)]$. Then
 - (i) $T \vdash \xi$ iff $N \models \Phi \prec \Psi$ iff $N \models \xi$;
 - (ii) $T \vdash \neg\xi$ iff $N \models \Psi \preceq \Phi$.
- (2) More generally, let for $i = 1, 2$, $T_i \supseteq I\Sigma_1$, let Pr_i be the proof predicate based on a fixed Δ_1 definition of T_i . Let

$$I\Sigma_1 \vdash \xi \equiv [(Pr_1(\neg\xi) \vee Pr_2(\neg\xi) \vee \Phi) \prec (Pr_1(\xi) \vee Pr_2(\xi) \vee \Psi)].$$

Then

- (i) $T_1 \vdash \xi$ iff $T_2 \vdash \xi$ iff $N \models \Phi \prec \Psi$ iff $N \models \xi$;
- (ii) $T_1 \vdash \neg\xi$ iff $T_2 \vdash \neg\xi$ iff $N \models \Psi \preceq \Phi$.

4.30 Lindström's Fixed Point Theorem. (Let T be as above.)

- (1) Let $\chi(y)$ be Σ_n and let $I\Sigma_1 \vdash \xi \equiv Pr_{\Sigma_n}(\neg\xi) \prec^\# (\exists y)\neg\chi(y)$. Then
 - (i) for each m , $(T + \xi) \vdash \chi(m)$,
 - (ii) for each $I\Sigma_1 \subseteq T_0 \subseteq T$ and each Π_n -sentence π , $(T_0 + \xi) \vdash \pi$ implies $T_0 + \{\chi(m) \mid m\} \vdash \pi$.
- (2) Let $\chi(y)$ be Π_n and let $I\Sigma_1 \vdash \xi \equiv (\exists y)\neg\chi(y) \prec^\# Pr_{\Pi_n}(\xi)$. Then
 - (i) for each m , $(T + \neg\xi) \vdash \chi(m)$,
 - (ii) for each $I\Sigma_1 \subseteq T_0 \subseteq T$ and each Σ_n -sentence β , $(T_0 + \neg\xi) \vdash \beta$ implies $T_0 + \{\chi(m) \mid m\} \vdash \beta$.

4.31 Remark. From these fixed point theorems we derive our three main results (4.25–4.27) as well as other results of independent interest (in particular, Theorem 4.34 claiming that each Σ_1 set has a sound Σ_1 numeration in T ; this was promised in 1.24). An elaboration follows.

4.32 (Proof of Smoryński's Fixed Point Theorem 4.29). We are going to prove part (2) of the theorem, (1) being a particular case. Recall that we have

two consistent theories $T_i \subseteq I\Sigma_1$ ($i = 1, 2$) and for both T_i we have a fixed definition of T_i which is Δ_1 in $I\Sigma_1$ (i.e. binumerates T_i in $I\Sigma_1$). Our diagonal formula ξ satisfies

$$I\Sigma_1 \vdash \xi \equiv [(Pr_1(\neg\xi) \vee Pr_2(\neg\xi) \vee \Phi) \prec (Pr_1(\xi) \vee Pr_2(\xi) \vee \Psi)]$$

where Φ, Ψ are fixed Σ_1 formulas. Thus ξ is Σ_1 in $I\Sigma_1$.

(1) First observe that if any of the formulas $\xi, Pr_1(\xi), Pr_2(\xi), Pr_1(\neg\xi), Pr_2(\neg\xi), \Phi, \Psi$ is true in N (i.e. has a standard witness d) then ξ becomes Δ_1 in $I\Sigma_1$ since the existential quantifier in the witness comparison formula can be bounded by $\bar{d}$.

(2) Thus we get

$$\begin{aligned} T_i \vdash \xi &\Rightarrow N \models \xi \Rightarrow T_j \vdash \xi, \\ T_i \vdash \neg\xi &\Rightarrow (N \models \neg\xi \text{ and } T_i \vdash \neg\xi) \end{aligned}$$

($i = 1, 2$) – note that $T_i \vdash \xi$ iff $N \models Pr_i(\xi)$.

(3) Furthermore, $T_i \vdash \xi$ implies $N \models \Phi \prec \Psi$ and similarly, $T_i \vdash \neg\xi$ implies $N \models \Psi \prec \Phi$. We prove the former claim: Let d be a T_i -proof of ξ ; thus it is both true and T_i -provable that there is an $y \leq \bar{d}$ which witnesses the antecedent $(Pr_1(\neg\xi) \vee Pr_2(\neg\xi) \vee \Phi)$ and $(\forall z \leq y)(z$ does not witness the succedent). Let $e \in N$ be such a y . But e cannot witness $Pr_j(\neg\xi)$ since this would make T_1 or T_2 contradictory; thus e witnesses Φ and hence $\Phi \prec \Psi$. We have $N \models \Phi \prec \Psi$.

(4) Conversely, $N \models \Phi \prec \Psi$ implies $T_i \vdash \xi$ since if $N \models \Phi \prec \Psi$ then $N \models \Phi$, which makes $\xi \Delta_1$ in $I\Sigma_1$ and hence decidable (provable or refutable) in T_i . But T_i cannot refute ξ since, by (3), this would imply $N \models \Psi \prec \Phi$. Similarly, $N \models \Psi \prec \Phi$ implies $T_i \vdash \neg\xi$. This gives all we need. $\square$

The theorem can be parametrized (and again called Sheperdson-Smoryński's fixed point theorem):

4.33 Theorem. Let $\varphi(x, y), \psi(x, y)$ be Σ_0 -formulas (or: formulas Δ_1 in $I\Sigma_1$), put $\Phi(x) \equiv (\exists y)\varphi(x, y), \Psi(x) \equiv (\exists y)\psi(x, y)$.

(1) Let T be as in 4.29 (1). Assume that for each k the formula $\xi(k)$ satisfies the following:

$$T \vdash \xi(\bar{k}) \equiv Pr(\neg\xi(\bar{k})) \vee \Phi(\bar{k}) \prec Pr(\xi(\bar{k})) \vee \Psi(\bar{k}).$$

Then

$$\begin{aligned} T \vdash \xi(\bar{k}) &\text{ iff } N \models \Phi(\bar{k}) \prec \Psi(\bar{k}) \text{ iff } N \models \xi(\bar{k}); \\ T \vdash \neg\xi(\bar{k}) &\text{ iff } N \models \Psi(\bar{k}) \prec \Phi(\bar{k}). \end{aligned}$$

(2) More generally, let T_i, Pr_i be as in 4.29 (2) (i.e. as in 4.32) and let for each k , $I\Sigma_1$ prove the following:

$$\xi(\bar{k}) \equiv (Pr_1(\neg\xi(\bar{k})) \vee Pr_2(\neg\xi(\bar{k})) \vee \Phi) \prec (Pr_1(\xi(\bar{k})) \vee Pr_2(\xi(\bar{k})) \vee \Psi).$$

Then

$$\begin{aligned} T_1 \vdash \xi(\bar{k}) &\text{ iff } T_2 \vdash \xi(\bar{k}) \text{ iff } N \models \xi(\bar{k}) \text{ iff } N \models \Phi(\bar{k}) \prec \Psi(\bar{k}), \\ T_1 \vdash \neg\xi(\bar{k}) &\text{ iff } T_2 \vdash \neg\xi(\bar{k}) \text{ iff } N \models \Psi(\bar{k}) \preceq \Phi(\bar{k}). \end{aligned}$$

The proof is fully analogous to 4.32; the existence of ξ is obvious. Part (1) of the parametric version will now be used to get several important consequences.

4.34 Theorem. If X, Y are disjoint Σ_1 sets then there is a Σ_1 -formula $\xi(x)$ such that, for each k ,

$$\begin{aligned} k \in X &\text{ iff } T \vdash \xi(\bar{k}) \text{ iff } N \models \xi(\bar{k}), \\ k \in Y &\text{ iff } T \vdash \neg\xi(\bar{k}). \end{aligned}$$

Proof. Take Σ_1 -definitions of X and Y for $\Phi(x)$ and $\Psi(x)$. □

4.35 Corollary. (1) Each Σ_1 set X has a sound numeration in T , i.e. a formula which both defines X and numerates X in T .

(2) Each Σ_1 set has a Π_1 numeration in T .

4.36 Theorem (= 4.26 (1)). There is no Δ_1 set X such that $Th \subseteq X \subseteq NRef$.

Proof. By 4.34, let ξ be such that $k \in X$ iff $T \vdash \xi(\bar{k})$ and $k \notin X$ iff $T \vdash \neg\xi(\bar{k})$; let $T \vdash \varphi \equiv \neg\xi(\bar{\varphi})$. Clearly, $\varphi \in X$ iff $\varphi \notin X$, a contradiction. □

4.37 Theorem. For each Π_1 set X there is a formula $\xi(x)$ such that, for each k ,

- (1) $k \in X$ iff $\xi(\bar{k})$ is neither Σ_1 -conservative nor Π_1 -conservative over T ,
- (2) $k \notin X$ iff $T \vdash \xi(\bar{k})$.

Proof. Observe that the formula $\xi(x)$ constructed for the Σ_1 -set $-X$ as in 4.29 is Σ_1 , thus Σ_1 -nonconservative whenever unprovable. Its Π_1 -nonconservativeness is proved similarly as the Π_1 -nonconservativeness of Rosser's formula. □

4.38 Corollary (= 4.26 (2)). There is no Π_1 -set X that $Th \subseteq X \subseteq Consv(\Sigma_1) \cup Consv(\Pi_1)$.

Proof. Assume we have such an X and take the corresponding formula $\xi(x)$ from 4.37. Let $T \vdash \varphi \equiv \xi(\bar{\varphi})$; if $\varphi \notin X$ then $T \vdash \xi(\bar{\varphi})$, thus $T \vdash \varphi$ and $\varphi \in X$, a contradiction. On the other hand, if $\varphi \in X$ then $\xi(\bar{\varphi})$ (and hence φ) is Σ_1 -nonconservative as well as Π_1 -nonconservative, i.e. $\varphi \notin X$, a contradiction. $\square$

We shall need the following corollary of Sheperdson-Smoryński's fixed point theorem.

4.39 Theorem. Let $T_1, T_2 \supseteq I\Sigma_1$ be Δ_1 and consistent, let X, Y be disjoint Σ_1 -sets. Then there is a Σ_1 -formula $\xi(x)$ such that, for each k ,

$$\begin{aligned} k \in X &\text{ iff } T_1 \vdash \xi(\bar{k}) \text{ iff } T_2 \vdash \xi(\bar{k}) \text{ iff } N \models \xi(\bar{k}), \\ k \in Y &\text{ iff } T_1 \vdash \neg \xi(\bar{k}) \text{ iff } T_2 \vdash \neg \xi(\bar{k}). \end{aligned}$$

In particular, $\xi(x)$ is a sound numeration of X both in T_1 and in T_2 .

Proof evident. $\square$

We now turn to proofs from true Γ -formulas (cf. 4.23). We write $\text{Proof}_\Gamma^\bullet(z, x)$ instead of $\text{Proof}_{T, \Gamma}^\bullet(z, x)$.

4.40 Lemma. For each φ and d ,

$$T \vdash \text{Proof}_\Gamma^\bullet(\bar{d}, \bar{\varphi}) \rightarrow \varphi;$$

consequently if for some d , T proves that $\bar{d}$ is a T -proof $^\bullet$ of $\bar{\varphi}$ from true Γ -formulas $^\bullet$ then T proves φ .

Proof. Let d and φ be given: assume that d is a sequence $\varphi_1, \dots, \varphi_n$ of T -formulas which is a $(T + \psi)$ -proof of φ for some $\psi \in \Gamma$, $\psi \leq d$. (If this is not the case then $T \vdash \neg \text{Proof}_\Gamma^\bullet(\bar{d}, \bar{\varphi})$ and we are done). Let us work in T .

Assume $\text{Proof}_\Gamma^\bullet(\bar{d}, \bar{\varphi})$. Then for some $y \in \Gamma$; $y \leq \bar{d}$, we have $\text{Proof}_{\Gamma+x}^\bullet(\bar{d}, \bar{\varphi})$ and $\text{Tr}_\Gamma(y)$, i.e. for some $\psi \leq d$ we have $\text{Tr}_\Gamma(\bar{\psi})$ and $\text{Proof}_{\Gamma+\bar{\varphi}}^\bullet(\bar{d}, \bar{\psi})$. Thus we get ψ . Having this we successively prove $\varphi_1, \varphi_2, \dots, \varphi_n$, hence φ . $\square$

4.41. We now prove Lindström's fixed point theorem (4.30).

- (1) Let $I\Sigma_1 \vdash \xi \equiv \text{Pr}_{\Sigma_n}(\neg \bar{\xi}) \prec^\# (\exists y) \neg \chi(y)$, $\chi \in \Sigma_n$;
recall the convention 4.23. Let $I\Sigma_1 \subseteq T_0 \subseteq T$. We shall prove the following:
 - (a) $T + \xi \vdash \chi(\bar{m})$,
 - (b) $T_0 + \xi$ is Π_n -conservative for $T_0 + \{\chi(\bar{m}) \mid m\}$.

Let us work in $T + \xi + \neg \chi(\bar{m})$; then m is a witness for the succedent of ξ , thus $(\exists y < \bar{m}) \text{Proof}_{\Sigma_n}^\bullet(y, \neg \bar{\xi})$; this gives $\neg \bar{\xi}$ by 4.40. We see that $(T + \xi + \neg \chi(\bar{m}))$ is contradictory. This proves (a).

Now assume $I\Sigma_1 \subseteq T_0 \subseteq T$ and let π be a Π_n -formula such that $(T_0 + \xi) \vdash \pi$, thus $(T_0 + \neg\pi) \vdash \neg\xi$; let d be a proof of $\neg\xi$ from the natural Σ_n equivalent δ of $\neg\pi$. Let us work in $T_0 + \{\chi(\bar{m}) \mid m\} + \neg\xi$, we want to prove π . Assume the contrary, i.e. δ ; then $Tr(\delta)$, i.e. $Proof_{\Sigma_n}^*(\bar{d}, \neg\xi)$, hence d is a witness of the antecedent of ξ . This gives $(\exists y \leq \bar{d}) \neg\chi(y)$, i.e. $\bigvee_{k \leq \bar{d}} \neg\chi(\bar{m})$, which is a contradiction in our theory. This proves (1).

(2) Now assume $I\Sigma_1 \vdash \xi \equiv (\exists y) \neg\chi(y) \prec^\# Pr_{\Pi_n}(\xi)$, $\chi \in \Pi_n$.

Let T_0 be as above. We shall prove

(a) $(\forall m)(T + \neg\xi \vdash \chi(\bar{m}))$,

(b) $(T_0 + \neg\xi)$ is Σ_n -conservative for $T_0 + \{\chi(m) \mid m\}$.

Let us work in $T_0 + \neg\xi + \chi(\bar{m})$. Then $\bar{m}$ witnesses the antecedent of ξ , and $\neg\xi$ implies $(\exists y \leq \bar{m}) Proof_{\Pi_n}^*(\bar{\xi}, y)$, hence ξ by 4.40, a contradiction. This proves (a).

Now assume $I\Sigma_1 \subseteq T_0 \subseteq T$ and let π be a Π_n -formula such that $(T_0 + \xi) \vdash \pi$, thus $(T_0 + \neg\pi) \vdash \neg\xi$; let d be a proof of $\neg\xi$ from the natural Σ_n equivalent δ of $\neg\pi$. Let us work in $T_0 + \{\chi(\bar{m}) \mid m\} + \neg\xi$, we want to prove π . Assume the contrary, i.e. δ ; then $Tr(\delta)$, i.e. $Proof_{\Sigma_n}^*(\bar{d}, \neg\xi)$, hence d is a witness of the antecedent of ξ . This gives $(\exists y \leq \bar{d}) \neg\chi(y)$, i.e. $\bigvee_{k \leq \bar{d}} \neg\chi(\bar{m})$, which is a contradiction in our theory. This proves (1).

(2) Now assume $I\Sigma_1 \vdash \xi \equiv (\exists y) \neg\chi(y) \prec^\# Pr_{\Pi_n}(\bar{\xi})$, $\chi \in \Pi_n$. Let T_0 be as above. We shall prove

(a) $(\forall m)(T + \neg\xi \vdash \chi(\bar{m}))$,

(b) $(T_0 + \neg\xi)$ is Σ_n -conservative for $T_0 + \{\chi(m) \mid m\}$.

Let us work in $T_0 + \neg\xi + \chi(\bar{m})$. Then $\bar{m}$ witnesses the antecedent of ξ , and $\neg\xi$ implies $(\exists y \leq \bar{m}) Proof_{\Pi_n}^*(\bar{\xi}, y)$, hence ξ by 4.40, a contradiction. This proves (a).

Now let $\delta \in \Sigma_n$ be such that $(T_0 + \neg\xi) \vdash \delta$, hence $(T_0 + \neg\delta) \vdash \xi$; let d be a corresponding proof. Work in $T_0 + \{\chi(\bar{m}) \mid m\} + \xi$; it suffices to prove δ . Assume the contrary, i.e. $Tr(\neg\delta)$; then $Proof_{\Pi_n}^*(\bar{d}, \bar{\xi})$, i.e. ξ gives $(\exists y \leq \bar{d}) \neg\chi(y)$, a contradiction. This completes our proof. $\square$

4.42 Generalizing Lindström's Fixed Point Theorem. There are two ways of generalization: first, we may let both the antecedent and the succedent depend on ξ ; second, we may parametrize the whole formula. We obtain the following four cases (T , Pr_{Σ_n} , Pr_{Π_n} as above, $I\Sigma_1 \subseteq T_0 \subseteq T$, $\chi(x, y)$ is Σ_n in (1), (2) and is Π_n in (3), (4).

(1)
$$I\Sigma_1 \vdash \xi \equiv Pr_{\Sigma_n}^*(\neg\bar{\xi}) \prec^\# (\exists y) \neg\chi(\bar{\xi}, y).$$

Properties: ξ is Σ_n in $I\Sigma_1$; $(T + \xi) \vdash \chi(\bar{\xi}, \bar{m})$ for all m ; $(T_0 + \xi)$ is Π_n -conservative for $T_0 + \{\chi(\bar{\xi}, \bar{m}) \mid m\}$.

(2)
$$I\Sigma_1 \vdash \xi(\bar{k}) \equiv Pr_{\Sigma_n}^*(\neg\bar{\xi}(\bar{k})) \prec^\# (\exists y) \neg\chi(\bar{k}, y).$$

Properties: $\xi(x)$ is Σ_n in $I\Sigma_1$; for all k, m , $(T + \xi(\bar{k}))\chi(\bar{k}, \bar{m})$; for all k , $(T_0 + \xi(\bar{k}))$ is Π_n -conservative for $T + \{\chi(\bar{k}, \bar{m}) \mid m\}$.

$$(3) \quad I\Sigma_1 \vdash \xi \equiv (\exists y)\neg\chi(\neg\xi, y) \prec^\# Pr_{\Pi_n}^\bullet(\bar{\xi}).$$

Properties: $\neg\xi$ is Π_n in $I\Sigma_1$; for all m , $(T + \neg\xi) \vdash \chi(\neg\xi, \bar{m})$; $(T_0 + \neg\xi)$ is Σ_n -conservative for $T_0 + \{\chi(\neg\xi, \bar{m}) \mid m\}$.

$$(4) \quad I\Sigma_1 \vdash \xi(\bar{k}) \equiv (\exists y)\neg\chi(\bar{k}, y) \prec^\# Pr_{\Pi_n}^\bullet(\bar{\xi}(\bar{k})).$$

Properties: $\neg\xi(x)$ is Π_n in $I\Sigma_1$; for all k, m , $(T + \neg\xi(\bar{k})) \vdash \chi(\bar{k}, \bar{m})$; $(T_0 + \neg\xi(\bar{k}))$ is Σ_n -conservative for $T_0 + \{\chi(\bar{k}, \bar{m}) \mid m\}$.

Proofs are obvious modifications of 4.41. □

4.43 Theorem. Let Γ be Σ_n or Π_n , $n \geq 1$. For each $X \in \Sigma_1$ there is a Γ -formula ξ such that, for each k ,

$$k \in X \text{ iff } T \vdash \neg\xi(\bar{k}),$$

$k \in X$ iff $\xi(\bar{k})$ is hereditarily $\tilde{\Gamma}$ -conservative over T . ($\tilde{\Gamma}$ is the dual of Γ).

Proof. In the preceding theorem, let $(\exists y)\neg\chi(x, y)$ define X . For $\Gamma = \Sigma_n$ use (2). Then: if $k \in X$ and m is a witness for $(\exists y)\neg\chi(\bar{k}, y)$ then

$$T + \xi(\bar{k}) \vdash \chi(\bar{k}, \bar{m}) \& \neg\chi(\bar{k}, \bar{m});$$

if $k \notin X$, i.e. for each m $I\Sigma_1 \vdash \chi(\bar{k}, \bar{m})$, then $T + \xi(\bar{k})$ is hereditarily Π_n -conservative over T .

Similarly for $\Gamma = \Pi_n$ use 4.42 (4). □

4.44 Corollary (= 4.26 (3)). There is no $X \in \Sigma_1$ such that

$$hConsv(\Gamma) \subseteq X \subseteq NRef.$$

Proof. Assume we have such an X and in the preceding theorem (with Γ and $\tilde{\Gamma}$ interchanged) let φ be such that $I\Sigma_1 \vdash \varphi \equiv \xi(\bar{\varphi})$. We show $\varphi \notin X$; but then 4.43 gives $\varphi \in hConsv(\Gamma)$, in contradiction to the inclusion assumed.

Thus let $\varphi \in X$; then $\varphi \in NRef$ (since $X \subseteq NRef$), but $T \vdash \neg\xi(\bar{\varphi})$, thus $T \vdash \neg\varphi$ (since $\varphi \in X$). A contradiction. □

4.45. We prove theorem 4.25 by inspecting the respective self-referential formulas.

(1) In 4.42 (1) let $(\exists y)\neg\chi(x, y)$ define the empty set, i.e. $I\Sigma_1 \vdash \chi(x, y) \equiv \overline{0} = \overline{0}$. Then $I\Sigma_1 \vdash \xi \equiv Pr_{\Sigma_n}(\neg\xi)$ (thus ξ says: my negation is provable in T from a true Σ_n -formula; in other words: I imply in T a false Π_n -formula), $I\Sigma_1 \vdash \chi(\bar{\xi}, \bar{k})$ for each k , thus ξ is hereditarily Π_n -conservative over T and clearly ξ is unprovable in T provided T is Σ_n -sound.

Lemma. Assume $T \vdash Pr(x) \rightarrow \Delta(x) \rightarrow Pr_{\Gamma}(x)$, and $T \vdash Pr(x) \rightarrow \nabla(x) \rightarrow Pr_{\Lambda}(x)$, $T \vdash \xi \equiv \Delta(\neg\xi) \prec^{\#} \nabla(\xi)$. Then neither ξ nor $\neg\xi$ is T -provable.

A standard proof (using 4.40) is left to the reader as an exercise.

(2) In 4.42 (1) put $\chi(x, y) \equiv \neg Proof^{\bullet}(y, x)$, i.e. $(\exists y)\neg\chi(x, y)$ is equivalent to $Pr^{\bullet}(x)$ and

$$I\Sigma_1 \vdash \xi \equiv Pr_{\Sigma_n}^{\bullet}(\neg\xi) \prec^{\#} Pr^{\bullet}(\bar{\xi}).$$

Then $I\Sigma_1 \vdash \chi(\bar{\xi}, \bar{m})$ for each m , $\neg\xi$ is unprovable and hereditarily Σ_n -conservative over T .

(3) In 4.42 (3) let $\chi(x, y)$ be as above, thus

$$I\Sigma_1 \vdash \xi \equiv Pr^{\bullet}(\neg\xi) \prec^{\#} Pr_{\Pi_n}^{\bullet}(\bar{\xi}).$$

Then $I\Sigma_1 \vdash \chi(\neg\xi, \bar{m})$ for each m , $\neg\xi$ is unprovable and hereditarily Σ_n -conservative over T .

(4) In 4.42(1) take $\neg Proof^{\bullet}_{\Pi_n}(y, x)$ for $\chi(x, y)$ and, at the same time, in 4.42 (3) take $\neg Proof^{\bullet}_{\Sigma_n}(y, x)$ for $\chi(x, y)$. In either case we get

$$I\Sigma_1 \vdash \xi \equiv Pr_{\Sigma_n}(\neg\xi) \prec^{\#} Pr_{\Pi_n}(\bar{\xi}).$$

Thus $T + \xi \vdash \neg Proof_{\Pi_n}(\bar{m}, \bar{\xi})$ by 4.42 (1), $T + \neg\xi \vdash \neg Proof^{\bullet}_{\Pi_n}(\bar{m}, \bar{\xi})$ by 4.40, thus 4.42 (1) gives Π_n -conservativity of ξ over T . (Nothing is claimed on hereditary conservativity: for a general T_0 between $I\Sigma_1$ and T we cannot guarantee $T_0 \vdash \neg Proof^{\bullet}_{\Pi_n}(\bar{m}, \bar{\xi})$.)

Similarly, 4.42 (3) and 4.40 gives $T \vdash \neg Proof^{\bullet}_{\Sigma_n}(\bar{m}, \bar{\xi})$ for each m and consequently $\neg\xi$ is Σ_n -conservative over T .

4.46 Theorem. Let T be Σ_n or Π_n , $n \geq 2$. Let X be Π_2 . Then there is a Γ -formula $\xi(x)$ such that, for each k ,

$$k \in X \equiv \xi(\bar{k}) \text{ is hereditarily } \tilde{\Gamma}\text{-conservative,}$$

$$k \notin X \equiv \xi(\bar{k}) \text{ is } \Sigma_1\text{-nonconservative.}$$

The same is true for Σ_1 replaced by Π_1 (with another ξ).

Proof. Let X be $\{k \mid (\forall m)R(k, m)\}$ where R is Σ_1 . By 4.39, let $\rho(x, y)$ be a Σ_1 -numeration (Π_1 -numeration) of R both in $I\Sigma_1$ and in T . In 4.42 (2), (4) let $\chi(x, y)$ be $\rho(x, y)$.

If $\Gamma = \Sigma_n$ let $\xi(x)$ be the formula $\xi(x)$ from 4.42 (2); if $\Gamma = \Pi_n$ let $\xi(x)$ be the formula $\neg\xi(x)$ from 4.42 (4). Then $\xi(x)$ is Γ ; if $k \in X$ then $I\Sigma_1 \vdash \rho(\bar{k}, \bar{m})$ for all m , hence $\xi(\bar{k})$ is hereditarily $\tilde{\Gamma}$ -conservative over T . If $k \notin X$ then, for some m_0 , T does not prove $\rho(\bar{k}, \bar{m}_0)$; thus $\xi(\bar{k})$ is not Σ_1 -conservative (Π_1 -conservative) over T . $\square$

4.47 Theorem (= 4.26). Let $\Gamma = \Sigma_n$ or Π_n , $n \geq 2$. There is no $Y \in \Sigma_2$ such that $hConsv(\Gamma) \subseteq Y \subseteq Consv(\Sigma_1)$ or $hConsv(\Gamma) \subseteq Y \subseteq Consv(\Pi_1)$.

Proof. Assume that Y is such set. Then 4.46 (with $X = N - Y$) gives ($k \notin Y$ iff $\xi(\bar{k})$ is hereditarily Γ -conservative). Let $I\Sigma_1 \vdash \varphi \equiv \xi(\bar{\varphi})$; then $\varphi \in Y$ implies that φ is not Σ_1 -conservative but $Y \subseteq Consv(\Sigma_1)$ a contradiction. On the other hand, $\varphi \notin Y$ implies that φ is hereditarily Γ -conservative but $hConsv(\Gamma) \subseteq Y$, a contradiction. $\square$

4.48 Theorem (= 4.27). For each $\Gamma = \Sigma_n, \Pi_n$ ($n \geq 1$), both $Consv(\Gamma)$ and $hConsv(\Gamma)$ is Π_2 -complete.

Proof. Clearly, $Consv(\Gamma)$ is Π_2 ; to see that $hConsv(\Gamma)$ is Π_2 , observe that φ is hereditarily Γ -conservative over T iff for each finite $T_0 \subseteq T$ containing $I\Sigma_1$, φ is Γ -conservative over T_0 . For each Γ (including Σ_1 and Π_1), 4.46 gives a formula $\xi(x)$ such that

$$\begin{aligned} k \in X &\Rightarrow \xi(\bar{k}) \text{ is hereditarily } \Gamma\text{-conservative,} \\ k \notin X &\Rightarrow \xi(\bar{k}) \text{ is not } \Gamma\text{-conservative.} \end{aligned} \quad \square$$

4.49. Now we present Švejdar's proof of the fact (claimed in 4.5 (3)) that for a T not Σ_1 -sound, the negation of Rosser's formula is Σ_1 -nonconservative over T .

Let $\delta(x)$ be a Σ_0 formula such that $T \vdash (\exists x)\delta(x)$ but $N \models (\forall x)\neg\delta(x)$; let ρ be Rosser's formula, thus $T \vdash \rho \equiv (Pr^*(\neg\rho) \prec Pr^*(\bar{\rho}))$. We know the following: both ρ and $\neg\rho$ are unprovable in T ;

$$T + Con^* \vdash \neg Pr^*(\rho) \& \neg Pr^*(\neg\rho)$$

(this is just a formalized version of the preceding assertion). Let

$$T \vdash \varphi \equiv ((\exists x)\delta(x) \prec Pr^*(\overline{\rho \rightarrow \varphi})).$$

- (a) $\rho \rightarrow \varphi$ is T -unprovable. Indeed, if d is a T -proof of $\rho \rightarrow \varphi$ then $T \vdash \bigvee_{k \leq d} \delta(\bar{k})$; but for each k , $T \vdash \neg\delta(\bar{k})$.
- (b) $T + Con^* \vdash \varphi$. Indeed, $T \vdash (\exists x)\delta(x)$, thus $(T + \neg\varphi)$ proves $Pr^*(\overline{\rho \rightarrow \varphi}) \prec (\exists x)\delta(x)$. (This is a Σ_1 -formula, call it σ), thus $T + \neg\varphi \vdash Pr^*(\bar{\sigma})$.

- But $T \vdash Pr^*(\sigma \rightarrow \neg\varphi)$, thus $T + \neg\varphi \vdash Pr^*(\neg\varphi) \ \& \ Pr^*(\sigma \rightarrow \varphi)$, hence $T + \neg\varphi \vdash Pr^*(\neg\rho)$, which implies $T + \neg\varphi \vdash \neg Con^*$.
- (c) $T + \neg\rho \vdash \varphi \vee (Pr^*(\bar{\rho}) \preceq Pr^*(\neg\bar{\rho}))$. This because $(T + \neg\rho + Con^*)$ proves ρ and $(T + \neg\rho + \neg Con^*)$ proves $(Pr^*(\bar{\rho}) \preceq Pr^*(\neg\bar{\rho}))$.
- (d) But $\varphi \vee (Pr^*(\bar{\rho}) \preceq Pr^*(\neg\bar{\rho}))$ is unprovable in T . Assume T proves this disjunction; then $T + \rho$ would prove φ (since $T + \rho$ disproves the second disjunct); this contradicts (a). Thus $\varphi \vee (Pr^*(\bar{\rho}) \preceq Pr^*(\neg\bar{\rho}))$ is the desired Σ_1 formula.

(c) Applications, Mainly to Interpretability

The subsection has the following structure: main results are formulated in 4.50–4.59 and compare (or, better, contrast) (i) interpretability in PA with interpretability in its conservative second order extension ACA_0 (4.55) and (ii) interpretability with partial conservativity for finitely axiomatized theories $T \supseteq I\Sigma_1$ (4.56–4.57). Proofs are elaborated in 4.50–4.65; this part starts with Lindström's second fixed point theorem, which is rather technical but of independent interest. The proofs combine tricky self-reference, partial truth definitions and use of definable cuts. The rest of the subsection (4.66–4.69) contains some additional results on partial conservativity.

4.50 Definition. Put $Intp_T = \{\varphi \mid (T + \varphi) \text{ is interpretable in } T\}$. Recall that we say " φ is interpretable in T " instead of " $(T + \varphi)$ is interpretable in T ". Further recall Theorem 3.46 telling us that, for T sequential and with full induction, $Intp_T = Consv_T(\Pi_1)$; thus Theorem 4.27 gives immediately the following

4.51 Corollary. If T is sequential and with full induction, then $Intp_T$ is Π_2 -complete.

4.52 Lemma. If $T \supseteq I\Sigma_1$ is finitely axiomatized then $Intp_T$ is Σ_1 .

Proof. This is more or less evident; φ is interpretable in T iff there are definitions of all (finitely many) symbols of T (in the sense of the interpretation) in T and T -proofs of translations of all (finitely many) axioms of the theory $(T + \varphi)$. Using finite sequences we may write down a formula $Interp_T^*(x, y) \Delta_1$ in $I\Sigma_1$ such that

$$\varphi \in Intp_T \text{ iff } N \models (\exists y) Interp_T^*(\bar{\varphi}, y);$$

$Interp_T^*$ just says that y is a sequence consisting of all those finitely many definitions and proofs. (cf. 2.42). $\square$

4.53 Theorem. If $T \supseteq I\Sigma_1$ is finitely axiomatized, $n \geq 1$ and Γ is Σ_n or Π_n then $Consv_T(\Gamma) - Intp_T \neq \emptyset$; there is a $\tilde{\Gamma}$ -sentence φ which is (hereditarily) Γ -conservative but is not interpretable in T .

Proof. This follows directly by 4.26, since clearly $Intp_T \subseteq NRef_T$ and $Intp_T \in \Sigma_1$, i.e. $hConsv_T(\Gamma)$ is not a subset of $Intp_T$. Checking the proof of 4.26 (3) (i.e. 4.44) we can see that the sentence constructed there is $\tilde{\Gamma}$. $\square$

4.54 Discussion. Take ACA_0 for T and Π_1 for Γ . We get a Σ_1 formula φ that is Π_1 -conservative over ACA_0 but $(ACA_0 + \varphi)$ is not interpretable in ACA_0 . Since φ is Π_1 -conservative over ACA_0 and is Σ_1 , i.e. a formula in the language of PA , φ is Π_1 -conservative over PA and hence, by 2.40, is interpretable in PA . (Similarly for $T = GB$, i.e. Gödel-Bernays set theory and ZF -Zermels-Frankel set theory, using 3.46). Note that we can take the following self-referential formula for φ :

$$T \vdash \varphi \equiv Pr_{\Sigma_1}^*(\neg\varphi) \prec (\exists y) Interp^*(\varphi, y).$$

This is slightly simpler than the formula obtained directly from 4.53, i.e. from 4.26.

Compare the present result with 4.16 and summarize:

4.55 Corollary. (1) There is a Σ_1 -formula φ such that $(PA + \varphi)$ is interpretable in PA but $(ACA_0 + \varphi)$ is not interpretable in ACA_0 .

(2) There is a Π_1 -formula φ such that $(ACA_0 + \varphi)$ is interpretable in ACA_0 but $(PA + \varphi)$ is not interpretable in PA .

(Similarly for ZF and GB instead of PA and ACA_0 .)

4.56 Remark. We shall analyze the present situation more deeply; we shall present a classification of independent Σ_1 -sentences φ (over an arbitrary consistent finitely axiomatized sequential theory $T \supseteq I\Sigma_1$) – with respect to the following questions. (i) is φ interpretable in T ? (ii) is $\neg\varphi$ interpretable in T ? (iii) is φ Π_1 -conservative over T ? We shall again get corollaries for $T = ACA_0$, GB and their relation to PA and ZF respectively.

Thus throughout the subsection, T is a consistent finitely axiomatized sequential theory containing $I\Sigma_1$. The question (i), (ii), (iii) admit eight possible combinations of answers and give eight possible types of independent Σ_1 -sentences. Till now, we have got some examples; e.g. for $\neg Con^*$ the answers are (yes, yes, no) and for the H -Rosser formula $H\rho$ (Rosser with respect to Herbrand proofs) the answers are (no, yes, yes). We are going to prove the following.

4.57 Theorem. There are independent Σ_1 -sentences of all eight possible types.

The theorem is an immediate consequence of Theorem 4.64 below. We shall construct examples of sentences of all eight types in a rather uniform way, using substantially results from Sect. 3.

4.58 Corollary. Take again $T = ACA_0$; recall that for each PA -sentence φ , φ is Π_1 -conservative over PA iff φ is Π_1 -conservative over ACA_0 iff $(PA + \varphi)$ is interpretable in PA . Consequently, if σ is an independent Σ_1 -sentence then $(PA + \neg\sigma)$ is never interpretable in PA . Hence we have eight possible types of dependent Σ_1 -sentences according to the following questions:

- is $(PA + \sigma)$ interpretable in PA ?
- is $(ACA_0 + \sigma)$ interpretable in ACA_0 ?
- is $(ACA_0 + \neg\sigma)$ interpretable in ACA_0 ?

Similarly for GB and ZF .

4.59 Remark. (1) We leave open the question which possibilities we have for Σ_1 (non)conservativity of $\neg\sigma$ in the case of T being Σ_1 -ill; in combination with the eight types above there at most 16 possible types. The reader may investigate this as an exercise.

(2) This is the end of the statement of results of the present subsection. The rest contains an elaboration, including Lindström's second fixed point theorem.

*

4.60 Lindström's Second Fixed Point Theorem. Let $T \supseteq I\Sigma_1$ and let $m, n \geq 0$. Let $\varphi(x, y)$ be a Σ_n -formula and $\theta(x, y)$ a Π_k -formula. Let ξ satisfy

$$T \vdash \xi \equiv [Pr_{\Sigma_n}^*(\neg\xi) \vee (\exists y)\varphi(\neg\xi, y) \prec^\# Pr_{\Pi_k}^*(\xi) \vee (\exists y)\theta(\xi, y)]$$

(where $Pr_{\Sigma_n}^*, Pr_{\Pi_k}^*$ are as in 4.23). Then

- (1) $(T + \xi) \vdash (\exists y \leq \bar{m})\theta(\xi, y) \rightarrow (\exists y \leq \bar{m})\varphi(\neg\xi, y)$ for each m ,
- (2) $(T + \neg\xi) \vdash (\exists y \leq \bar{m})\varphi(\neg\xi, y) \rightarrow (\exists y \leq \bar{m})\theta(\xi, y)$ for each m ,
- (3) $(T + \xi)$ is Π_n -conservative for $T + \{\neg\theta(\xi, \bar{m}) \mid m\}$,
- (4) $(T + \neg\xi)$ is Σ_k -conservative for $T + \{\neg\varphi(\neg\xi, \bar{m}) \mid m\}$.

Furthermore, ξ is $\Sigma_{\max(k, n, 1)}$ in T .

The assertion remains valid if Pr^* is replaced by HPr^* (Herbrand provability).

Proof. (1) Let us work in $(T + \xi + (\exists y \leq \bar{m})\theta(\xi, y))$. The succedent of ξ has a witness $\leq \bar{m}$; this implies in the usual way that the antecedent of ξ has also such a witness; thus $(\exists y < \bar{m})(Proof_{\Sigma_n}^*(\neg\xi) \vee \varphi(\neg\xi, y))$. But this y is not a witness for $Pr_{\Sigma_n}^*(\neg\xi)$ (otherwise we would get ξ by 4.40 or by the variant of 4.40 using HPr^*). Thus we get $(\exists y \leq \bar{m})\varphi(\neg\xi, y)$.

(2) is proved analogously.

(3) Assume $(T + \xi) \vdash \pi$ where π is Π_n , let d be a T -proof of $(\neg\pi \rightarrow \neg\xi)$. Let us work in $(T + \{\neg\theta(\bar{\xi}, \bar{m}) \mid m\} + \neg\xi)$; we want to prove π . Assume $\neg\pi$, thus $Tr(\neg\pi)$, i.e. $\neg\pi$ is a true Σ_n -sentence; but then $Proof_{\Sigma_n}(\bar{d}, \neg\xi)$ and, since we assume $\neg\xi$, beneath $\bar{d}$ there is a witness for the succedent of ξ . But beneath $\bar{d}$ there is no witness for $(\exists n)\theta(\bar{\xi}, u)$ (since $\neg\theta(\bar{\xi}, \bar{0}), \neg\theta(\bar{\xi}, \bar{1}), \dots, \neg\theta(\bar{\xi}, \bar{d})$). Neither is there a witness for $Pr_{\Pi_k}^*(\bar{\xi})$ beneath $\bar{d}$ since this would imply ξ by 4.40. This gives a contradiction.

(4) similarly. $\square$

4.61 Remark. For $n = 0$ we may replace $Pr_{\Sigma_0}^*$ by Pr^* and/or allow φ to be Δ_1 (in $I\Sigma_1$, say) instead of being Σ_0 . Similarly for $k = 0$.

4.62 Corollary. In 4.60, let m, n be arbitrary but let φ, θ be Σ_1 -in- $I\Sigma_1$. (By 4.61, this does not exclude the possibility $m = 0$ and/or $n = 0$). Let X, Y be Σ_1 -sets defined by $(\exists y)\theta(x, y)$ and $(\exists y)\varphi(x, y)$ respectively and assume $X, Y \subseteq NRef$ (non-refutable formulas). Then the sentence ξ from 4.60 is Π_n -conservative, $\neg\xi$ is Σ_k -conservative (over T), $\xi \notin X$ and $(\neg\xi) \notin Y$.

Proof. Without loss of generality we may assume that, for each k , no m witnesses both $k \in X$ and $k \in Y$, i.e. $N \models (\forall x, y)(\neg\varphi(x, y) \vee \theta(x, y))$ (assume e.g. that all witnesses for X are even and all witnesses for Y are odd). Take ξ from 4.60, i.e.

$$I\Sigma_1 \vdash \xi \equiv Pr_{\Sigma_n}^*(\neg\bar{\xi}) \vee (\exists y)\varphi(\neg\bar{\xi}, y) \prec^\# Pr_{\Pi_k}^*(\bar{\xi}) \vee (\exists y)\theta(\bar{\xi}, y).$$

First show $\xi \notin X$ and $(\neg\xi) \notin Y$. Assume the contrary and let m be the least witness for $\xi \in X \vee (\neg\xi) \in Y$. Thus either N satisfies (and T proves)

$$\theta(\bar{\xi}, \bar{m}) \& (\forall x < \bar{m})\neg\theta(\bar{\xi}, x) \& (\forall x \leq \bar{m})\neg\varphi(\neg\bar{\xi}, x),$$

or N satisfies and T proves the formula resulting from the last one by interchanging $\theta, \bar{\xi}$ with $\varphi, \neg\bar{\xi}$ respectively. In the former case we get a contradiction in $(T + \xi)$ using 4.60 (1), which contradicts our assumption $X \subseteq NRef$; in the latter case 4.60 (2) gives a contradiction in $(T + \neg\xi)$, which contradicts $Y \subseteq NRef$. Thus we have $\xi \notin X$ and $\xi \notin Y$. This implies $T \vdash \neg\theta(\bar{\xi}, \bar{m})$ for each m and therefore 4.60 (3), (4) give the desired conservation. $\square$

4.63 Discussion and Definition. Now we shall formulate and prove a theorem implying immediately our Theorem 4.57. We shall deal with formulas $HPr^*(x)$, $HPr_{\Sigma_1}^*(x)$ (Herbrand provability in T ; Herbrand provability in T^* from a true Σ_1 -sentence) as well as with the formula $Intp^*(x) \equiv (\exists y)Interp^*(x, y)$ from 4.52 (interpretability of $(T^* + x)$ in T^*). We shall investigate self-referential Σ_1 -sentences ξ such that

$$I\Sigma_1 \vdash \xi \equiv \Delta(\neg\bar{\xi}) \prec \nabla(\bar{\xi}),$$

where the antecedent $\Delta(x)$ has one of the forms

$$HPr^\bullet(x), \quad HPr^\bullet_{\Sigma_1}(x), \quad HPr^\bullet(x) \vee Intp^\bullet(x), \quad HPr^\bullet_{\Sigma_1}(x) \vee Intp^\bullet(x),$$

and the succedent $\nabla(x)$ has one of the forms

$$HPr^\bullet(x), \quad HPr^\bullet(x) \vee Intp^\bullet(x).$$

This will give eight cases, which are examples of our eight types of independent Σ_1 -sentences. For some of the following assertions it is immaterial whether we use $HPr^\bullet$ or $Pr^\bullet$, but for some (using Theorem 3.20) it is not.

Observe that all eight cases are particular cases of 4.62 (with the convention 4.61 applied) for $n = 0$ or 1 , $k = 0$ and X, Y being either empty or $Intp_T$; we use Herbrand provability.

4.64 Theorem. (1) All cases give independent Σ_1 -sentences;

(2) ξ is interpretable iff $X = \emptyset$ (the succedent does not contain $Intp^\bullet$);

(3) $\neg\xi$ is interpretable iff $Y = \emptyset$ (the antecedent does not contain $Intp^\bullet$);

(4) ξ is Π_1 -conservative iff $n = 1$ (the antecedent contains $HPr^\bullet_{\Sigma_1}$).

Remark. This obviously implies Theorem 4.57.

Proof. 4.62 gives directly the following: ξ is independent (since ξ is Π_n -conservative and $\neg\xi$ is Σ_k -conservative); if the antecedent contains $HPr^\bullet_{\Sigma_1}$ then ξ is Π_1 -conservative; if $X = Intp_T$ then ξ is not interpretable; if $Y = Intp_T$ then $\neg\xi$ is not interpretable. Thus it remains to prove the converse implications in (2)–(4).

(2) Assume $X = \emptyset$; we prove $\xi \in Intp_T$. (Here it is vital to work with Herbrand provability.) We have $T \vdash \xi = \Delta(\neg\xi) \prec HPr^\bullet_\Gamma(\bar{\xi})$. It suffices to interpret $(T + \xi)$ in $(T + \neg\xi)$. By 3.20 there is a $(T + \neg\xi)$ -cut J such that $(T + \neg\xi) \vdash HCon^{\bullet J}(T^\bullet + \neg\xi)$. Let us work in $(T + \neg\xi)$. In J there is no H -proof[•] of $\bar{\xi}$ in $T^\bullet$. But $\neg\xi$ implies that beneath each witness for $\Delta(\neg\xi)$, in particular, beneath each H -proof[•] of $\neg\xi$ in $T^\bullet$ there is a H -proof[•] of $\bar{\xi}$ in $T^\bullet$, thus in J there is no H -proof[•] of $\neg\bar{\xi}$ in $T^\bullet$ and hence $HCon^{\bullet J}(T^\bullet + \bar{\xi})$. By 3.39, $(T + \bar{\xi})$ is interpretable $(T + \neg\xi)$.

(3) Assume $Y = \emptyset$; we prove $(\neg\xi) \in Intp_T$. We have $T \vdash \xi \equiv HPr^\bullet_\Gamma(\neg\xi) \prec \nabla(\bar{\xi})$ where $\Gamma = \Sigma_0$ or Σ_1 . It suffices to interpret $(T + \neg\xi)$ in $(T + \xi)$. Let us work in $(T + \xi)$. By 3.30 we have a cut J such that $(\forall u \in Tr(\Sigma_1)) HCon^{\bullet J}(T^\bullet + \bar{\xi} + u)$, i.e. J does not contain any H -proof[•] of $(\neg T^\bullet \vee \neg\bar{\xi} \vee u)$. But ξ implies that there is a H -proof[•] y of $\neg\bar{\xi}$ from T and from a true Σ_1 -formula u_0 (i.e. a H -proof[•] of $\neg T^\bullet \vee \neg u_0 \vee \neg\bar{\xi}$) such that beneath y there is no witness for $\nabla(\bar{\xi})$, in particular, no H -proof[•] of $\bar{\xi}$ from $T^\bullet$, hence we have $HCon^{\bullet J}(T' + \neg\bar{\xi})$; by 3.39, $(T + \neg\xi)$ is interpretable in $(T + \xi)$.

(4) Assume $n = 0$, i.e. $HPr^\bullet_{\Sigma_1}$ does not occur in ∇ . We prove that ξ is Π_1 -nonconservative. Clearly, ξ implies (in T) the formula saying “beneath

each witness for $\nabla(\bar{\xi})$ there is a witness for $\nabla(\neg\bar{\xi})$ (call it π) and under our assumptions π is Π_1 . It suffices to show that π is unprovable in T . This is done analogously to the corresponding proof concerning Rosser's formula. If π were T -provable we would have

$$T \vdash (\text{there exists a witness for } \nabla(\bar{\xi}) \rightarrow \xi)$$

(consider the least witness for $\nabla(\neg\bar{\xi})$, which exists by $I\Sigma_1$). Thus we would have

$$T \vdash \neg\xi \rightarrow (\nabla(\bar{\xi}) \text{ has no witness}),$$

$$T \vdash \neg\xi \rightarrow \neg HPr^*(\bar{\xi}),$$

$$T \vdash \neg\xi \rightarrow HCon^*(T^* + \neg\bar{\xi}),$$

$$T \vdash \neg\xi \rightarrow Con^*(T^* + \neg\bar{\xi})$$

(due to the provability of Herbrand's theorem in $I\Sigma_1$), which contradicts Gödel's second incompleteness theorem. This completes the proof of 4.66 and of 4.57. $\square$

*

To close the present section we add some few other applications of partial conservativity. $T \in \Delta_1$ is a consistent theory containing $I\Sigma_1$.

4.66 Theorem (speed up). Let f be a Δ_1 function mapping N into N , let $\Gamma = \Sigma_k$ or Π_k ($k \geq 1$). There is a φ Γ -conservative over T and such that, for each n , there is a proof $d > n$ of a Γ -sentence γ in $(T + \varphi)$ such that each T -proof of γ is bigger than $f(d)$.

Proof. Otherwise $Consv(\Gamma)$ would be Σ_2 , which contradicts 4.27. (Since a Π_2 -complete set is not Σ_2 .) $\square$

4.67 Definition and discussion. A sentence φ is a *self-prover* if $T \vdash \varphi \rightarrow Pr^*(\varphi)$. Clearly, each Σ_1 -sentence is a self-prover. We show that there are other self-provers as well. It is easily seen that for each sentence ψ , the formula $\psi \ \& \ Pr^*(\psi)$ is a self-prover; if φ is a self-prover then $T \vdash \varphi \equiv (\varphi \ \& \ Pr^*(\varphi))$.

4.68 Theorem. For each $\Gamma = \Sigma_n$ ($n \geq 1$) or $\Gamma = \Pi_m$ ($m \geq 2$) there is a self-prover which is exactly of complexity Γ .

Proof. By 4.26 (3), let ψ be such that $\neg\psi$ is Γ -conservative over T and not provable in $(T + \neg Con^*)$ (i.e. take $X = \{\alpha \mid (T + \neg Con) \vdash \alpha\}$). Let φ be $\psi \ \& \ Pr^*(\psi)$. By our assumption on Γ , φ is Γ in T . We prove that φ is not $\tilde{\Gamma}$ (dual class). Assume $T \vdash \varphi \equiv \sigma$, $\sigma \in \tilde{\Gamma}$. Then $T \vdash \varphi \rightarrow \neg\sigma$, hence $T \vdash \neg\psi \rightarrow \neg\sigma$, and since $\neg\psi$ is Γ -conservative we get $T \vdash \neg\sigma$, thus $T \vdash \neg\varphi$,

$T \vdash \neg(\psi \& Pr^*(\bar{\psi}))$, $T \vdash Pr^*(\bar{\psi}) \rightarrow \neg\psi$, $T \vdash \neg Con^* \rightarrow \neg\psi$, which contradicts our assumption. $\square$

4.69 Remark. Let Γ be Σ_n or Π_n , let Λ be Σ_m or Π_m , assume that Λ does not include Γ . Then there is a formula φ exactly Γ (not $\tilde{\Gamma}$ in T) and hereditarily exactly Λ -conservative (not $\tilde{\Lambda}$ -conservative). For example, let $\Gamma = \Sigma_8$, let Λ be Σ_5 ; let π be Π_5 , independent and hereditarily Σ_5 -conservative over T and let σ be Σ_8 , independent, hereditarily Π_8 -conservative over $(T + \pi)$. Then take $\varphi \equiv \pi \& \sigma$.

Chapter IV

Models of Fragments of Arithmetic

Introduction. The present chapter is devoted to the study of models of arithmetic, i.e. structures M for the language L of arithmetic such that $M \models T$, where T is either PA or a fragment of PA . Models are useful as means for showing unprovability of a formula φ in T (by exhibiting a model M for $T + \neg\varphi$) as well as for showing provability (by proving that $(T + \neg\varphi)$ has no model). In particular, we shall prove some conservation results by model-theoretical methods.

Model theory of arithmetic is a rather broad field and we shall not try to be exhaustive; our aim will be to present selected typical results and techniques. Section 1 (Some basic constructions) makes the reader familiar with non-standard models and some techniques of their construction; the main results are hierarchy results (concerning theories $I\Sigma_n$, $B\Sigma_n$, $P\Sigma_n$), theorems concerning existence of elementary extensions and the Paris-Friedman conservation result. Section 2 (Cuts in models of arithmetic with top) introduces and studies cuts (roughly: initial segments of models having no greatest element) and various kinds of cuts: in particular, k -extendable cuts and their relationship to models of $B\Sigma_{k+1}$. Furthermore, the section contains a proof of the theorem saying that each non-standard model of $I\Sigma_1$ is isomorphic to a proper cut of itself. Section 3 (Provably recursive functions and the method of indicators) presents two characterizations of a $I\Sigma_k$ -provably recursive function; a corollary is the fact that $I\Sigma_1$ -provably recursive functions coincide with primitive recursive functions. We rely heavily on the method of indicators, a formalization of ordinals in $I\Sigma_1$, k -extendable (or k -restrainable) cuts and the notion of α -large sets. Finally, there is a short Sect. 4, dealing with a formalization of some parts of model theory of fragments; we formalize in $I\Sigma_1$. In particular, we give in $I\Sigma_1$ a model-theoretical proof of Paris-Friedman's conservation theorem and get various corollaries, among them the fact that $I\Sigma_{k+1}$ proves the consistency of $B\Sigma_{k+1}^\bullet$.

Our main omission is the absence of the theory of recursively saturated models. We apologize for this and refer to bibliographical remarks for references. But we hope that the present chapter will give the reader satisfactory insight into the world of models of arithmetic and their possible uses.

1. Some Basic Constructions

(a) Preliminaries

1.1 Introduction. Our starting point is the *standard* model N , consisting of all natural numbers endowed with the usual (standard) successor, addition, multiplication, zero, equality and the less-than-or-equal-to relation. Since isomorphic models differ in an inessential manner, each model isomorphic to N is also called standard. We shall be interested in non-standard models. We shall investigate only *countable* models; thus in saying “model” we always mean “countable model”.

We shall often use Gödel's famous completeness theorem 0.12 and also the method of skolemization 0.14. We shall also make use of the fact that Matiyasevič's theorem is provable in IS_1 (cf. I.1.59). We shall construct new models as extensions of a given model (e.g. of N) or as submodels of a given (non-standard) model. In this section we shall deal with *definable ultrapower* as a method for construction of an extension of a given model and with a few methods of construction of *submodels* using *definable elements*. Definable ultrapower of N is one of the simplest examples of a non-standard model (and corresponds – mutatis mutandis – to the first historical construction of a non-standard model of PA due to Skolem). This is elaborated on in subsection (b). The methods of constructing submodels of a non-standard model are powerful enough to produce, for each theory T from the hierarchy of theories $IS_k, B\Sigma_{k+1}, P\Sigma_{k+1}$, a model of T which is not a model of any stronger theory in the hierarchy. For details see subsection (d). In (e) we generalize the construction of definable ultrapower and characterize models of $B\Sigma_{k+1}$ by the existence of some proper extension; as a by-product we obtain new models of some theories in our hierarchy that are not models of stronger theories. Subsection (f) contains two proofs of the theorem of Friedman and Paris saying that $B\Sigma_{k+1}$ is Π_{k+2} -conservative over IS_k ($k \geq 0$). Thus the reader finds here some techniques for construction of models, examples of models showing that the hierarchy of fragments is proper (and non-linear) and the above-mentioned conservation result. Most results of this section are due – mutatis mutandis – to Skolem, Gaifman, Friedman, Paris, Kirby, Wilkie, Lessan, Kaye and Dimitracopulos; see bibliographical notes for details.

1.2 Conventions. M, K, I, J vary over models for L ; N is reserved for the standard model. We use the same letter for a model and its domain if there is no danger of misunderstanding. A model M has its zero 0_M , successor S_M , addition $+_M$, multiplication $\cdot_M$, inequality $\leq_M$; we deal with models having absolute equality. Sometimes we shall omit the index M in $+_M$ etc. If $t(x_1, \dots, x_n)$ is a term L containing the variables displayed then t_M is the corresponding mapping of M^n into M , i.e., for each $a_1, \dots, a_n \in M$,

$t_M(a_1, \dots, a_n)$ is the value of t for $a_1, \dots, a_n$. We use Tarski's definition of satisfaction; if φ is closed then $M \models \varphi$ means that φ is true in M . If $\varphi(x_1, \dots, x_n)$ has the free variables displayed then $M \models \varphi(a_1, \dots, a_n)$ means that φ is satisfied by $a_1, \dots, a_n$ in M (which in turn is a condition concerning sequences of elements of M having the value a_i in the place corresponding to x_i). $M \subseteq K$ means that M is a submodel of K , i.e. the domain of M is a subset of the domain of K , $0_M = 0_K$ and $S_M, +_M, *_M, \leq_M$ are restrictions of $S_K, +_K, *_K, \leq_K$ respectively.

1.3 Some Definitions and Facts. (1) If $M \models I_{\text{open}}$ then $+_M$ and $*_M$ are associative and commutative, satisfy the distributive law for multiplication w.r.t. addition, $\leq_M$ is a discrete linear order with the least element 0_M , S_M associates to each element its bigger neighbour, addition is monotone and so is multiplication by a non-zero element; etc., cf. Chap. I, Sect. 1.

(2) Let $K, M \models I_{\text{open}}$. K is an *end-extension* of M (M is a *cut* in K , notation $M \subseteq_e K$) if for each $a \in M$ and each $b \in K - M$ we have $K \models a \leq b$, i.e. each element of $M \leq_K$ -precedes each element of $K - M$.

(3) Each $M \models I_{\text{open}}$ is an end-extension of (a copy of) the standard model:

Let $I = \{a \in M \mid (\exists n \in N)(M \models a = \bar{n})\}$; then clearly I is a cut in M (cf. I.1.6 (4)) and is isomorphic to N . Evidently, M is non-standard iff $I \neq M$, i.e. iff there is an $a \in M$ such that for each $n \in N$, $M \models a \neq \bar{n}$.

(4) If $M, K \models I_{\text{open}}$, $M \subseteq_e K$, $\varphi \in \Sigma_0$ and $a, \dots, b \in M$ then $M \models \varphi(a, \dots, b)$ iff $K \models \varphi(a, \dots, b)$ (obvious). Consequently, if $\varphi \in \Sigma_1$ and $a, \dots, b \in M$ then $M \models \varphi(a, \dots, b)$ implies $K \models \varphi(a, \dots, b)$.

(5) In particular, the set of all L -formulas is defined in N by a formula $\text{Form}^*(x)$ which is Δ_1 in $I\Sigma_1$. Thus φ is a formula iff $N \models \text{Form}^*(\bar{\varphi})$ iff $M \models \text{Form}^*(\bar{\varphi})$ for each $M \models I\Sigma_1$.

(6) We introduce the notion of a Σ_n -definable subset of a model $M \models I\Sigma_0$; $A \subseteq M$ is *parametrically Σ_n -definable in M* iff there is a Σ_n -formula φ and an element $c \in M$ such that $A = \{a \in M \mid M \models \varphi(a, c)\}$. Similarly for non-parametric definability and for definable subsets of M^n .

(7) Recall that in I.1.73–74 we defined satisfaction for Σ_n -formulas in $I\Sigma_1$ and in I.1.81 we defined Σ_n -definable sets and their codes in $I\Sigma_1$; we have predicates $\text{Code}_{\Sigma, n}$ and $\in_{\Sigma, n}$. The relation to definable subsets of models is as follows:

1.4 Claim. Let $M \models I\Sigma_1$ and $A \subseteq M$. The following is equivalent:

- (1) A is parametrically Σ_n -definable in M ;
- (2) there is a $c \in M$ such that $M \models (c \text{ is a } \Sigma_n\text{-formula}^* \text{ with one variable})$ and $A = \{a \in M \mid M \models \text{Sat}_{\Sigma, n}(c, \langle a \rangle)\}$
- (3) there is a $d \in M$ such that $M \models \Sigma_n^*\text{-set}^*(d)$ and $A = \{a \in M \mid M \models a \in_{\Sigma, n} d\}$. (Cf. I.1.78)

Proof. (1) $\Rightarrow$ (2). Let $A = \{a \mid M \models \varphi(a, p)\}$; $M \models \varphi(a, p)$ iff $M \models \text{Sat}_{\Sigma, n}(\bar{\varphi}, \langle a, p \rangle)$ iff $M \models \text{Sat}_{\Sigma, n}(c, \langle a \rangle)$ where c results from $\bar{\varphi}$ by substituting the p -th numeral for the second variable).

(2) $\Rightarrow$ (3). Let $A = \{a \mid M \models \text{Sat}_{\Sigma, n}(c, \langle a \rangle)\}$, then $M \models \Sigma_n^* \text{-set}^*(c)$ and $A = \{a \mid a \in_{\Sigma, n} c\}$.

(3) $\Rightarrow$ (1). If $A = \{a \in M \mid M \models a \in_{\Sigma, n} d\}$ then A is Σ_n -definable in M since $\in_{\Sigma, n}$ is a Σ_n -formula. $\square$

Example of a non-standard model. The existence of a non-standard-model follows trivially from compactness and completeness: let $Th(N)$ be the set of all sentences true in N . Let α be a new constant and put $T' = T \cup \{\bar{n} \leq \alpha \mid n \in N\}$. T' is consistent by compactness and therefore has a model M ; M is non-standard since if c is the meaning of α then $M \models c \geq \bar{n}$ for each $n \in N$. In the next subsection we shall *construct* a non-standard model in an apparently more transparent way.

(b) Definable Ultrapower of the Standard Model

1.6 Construction. Let F be the set of all definable mappings of N into N (as particular subsets of N^2). Let D be the set of all definable subsets of N . Clearly, D is a Boolean algebra, i.e. is closed under meet, union and complement. D is countable; let U be a non-trivial ultrafilter in D , i.e. we have the following:

$$\begin{aligned} A \subseteq B \text{ and } A \in U &\text{ implies } B \in U. \\ A, B \in U &\text{ implies } A \cap B \in U, \\ \text{for each } A \in D, &\text{ either } A \in U \text{ or } (N - A) \in U, \\ A \in U &\text{ implies that } A \text{ is infinite.} \end{aligned}$$

(Hint: let $A_0, A_1, A_2, \dots$ be the sequence of all (countably many) definable subsets of N ; define $D_0 = N$ and for each n , let D_n be $D_0 \cap \dots \cap D_{n-1} \cap A_n$ if this set is infinite, otherwise let $D_n = D_0 \cap \dots \cap D_{n-1} \cap (N - A_n)$. Let $A \in D$ iff $(\exists n)(D_n \subseteq A)$.)

Let, for $f, g \in F$, $f =_U g$ mean $\{i \in N \mid N \models f(i) = g(i)\} \in U$. (Observe that the last set is in D .) It is easy to show that $=_U$ is an equivalence, i.e. is reflexive, symmetric and transitive. Let $[f]_U$ be the equivalence class of f , i.e. $[f]_U = \{g \mid f =_U g\}$. Let $N^* = \{[f]_U \mid f \in F\}$. Put $f +_F g = h$ iff, for each $i \in N$, $f(i) + g(i) = h(i)$, similarly for $S, *$. Let $f \leq_U g$ mean $\{i \in N \mid f(i) \leq g(i)\} \in U$. Show that $=_U$ is a congruence w.r.t. the operations $S_F, +_F, *_F$; this means that we may define $[f]_U +_{N^*} [g]_U = [f +_F g]_U$ and similarly for S . Doing this we endow N^* with a structure for L (we put $0_N = k_0$ where $k_0(i) = 0$ for each i). This structure is denoted by N^* (or $N^*(U)$) and is called the *definable ultrapower* of N (given by U).

1.7 Theorem. (1) N^* is a structure for L elementarily equivalent to N ; thus N^* is a model of PA .

(2) The mapping associating with each $n \in N$ the element $[k_n]_U$ (where $k_n(i) = n$ for each i) is the unique isomorphic embedding of N onto a cut in N^* .

(3) N^* is non-standard; e.g. the diagonal d (such that $d(i) = i$ for each i) is a non-standard element.

The rest of the subsection consists of the proof of this theorem.

1.8 Lemma. (1) For each term $t(x_1, \dots, x_n)$ of L and all $f_1, \dots, f_n \in F$, let $t_F(f_1, \dots, f_n)$ be the function $f(i) = t_N(f_1(i), \dots, f_n(i))$. Then $t_{N^*}([f_1]_U, \dots, [f_n]_U) = [t_F(f_1, \dots, f_n)]_U$.

(2) For any terms $t(x_1, \dots, x_n)$, $s(x_1, \dots, x_n)$ and $f_1, \dots, f_n \in F$,

$$N^* \models t(f_1, \dots, f_n) = s(f_1, \dots, f_n)$$

iff

$$\{i \in N \mid N \models t(f_1(i), \dots, f_n(i)) = s(f_1(i), \dots, f_n(i))\} \in U.$$

The same replacing $=$ with $\leq$.

Proof. (1) by induction on the complexity of t . (2) follows immediately from (1). $\square$

1.9 Łoś's Lemma. For each L -formula $\varphi(x_1, \dots, x_n)$ and any $f_1, \dots, f_n \in F$,

$$N^* \models \varphi([f_1]_U, \dots, [f_n]_U) \text{ iff } \{i \in N \mid N \models \varphi(f_1(i), \dots, f_n(i))\} \in U.$$

Proof by induction on the complexity of φ ; we elaborate the induction step for quantification. Let φ be $(\exists y)\psi(x_1, \dots, x_n, y)$ and assume the assertion for ψ . If $N^* \models (\exists y)\psi([f_1]_U, \dots, [f_n]_U, [g]_U)$ then for some g , $N^* \models \varphi([f_1]_U, \dots, [f_n]_U, [g]_U)$, thus

$$\{i \mid N \models \varphi(f_1(i), \dots, f_n(i), g(i))\} \in U,$$

which implies $\{i \mid N \models (\exists u)\varphi(f_1(i), \dots, f_n(i), u)\} \in U$. Conversely, let $A = \{i \mid N \models (\exists u)\varphi(f_1(i), \dots, f_n(i), u)\} \in U$. Define g as follows:

$$\begin{aligned} \text{for } i \in A, g(i) &= \min_j N \models \varphi(f_1(i), \dots, f_n(i), j); \\ \text{for } i \notin A, g(i) &= 0. \end{aligned}$$

Since $f_1, \dots, f_n$ are definable, one easily sees that A is definable and so is g . Clearly,

$$A = \{i \mid N \models \varphi(f_1(i), \dots, f_n(i), g(i))\} \in U,$$

thus $N^* \models \psi([f_1]_U, \dots, [f_n]_U, [g]_U)$ and $N^* \models (\exists u)\psi([f_1]_U, \dots, [f_n]_U, u)$. $\square$

1.10 Corollary. 1.7 (1) follows.

1.11 Lemma. For each n , let $k_n(i) = n$ for each i ; then k_n is the value of the n -th numeral $\bar{n}$ in N .

Proof. Clearly, $[k_0]_U$ is the least element in N^* and $n < m$ implies $N^* \models [k_n]_U < [k_m]_U$. It is sufficient to show that if $N^* \models [f]_U \leq [k_n]_U$ then for some $m < n$, $[f]_U = [k_m]_U$. Let $A_m = \{i \mid f(i) = m\}$; clearly, $\{i \mid f(i) < n\} = A_0 \cup \dots \cup A_n \in U$ and therefore for an $m < n$ we have $A_m \in U$. Thus $f =_U k_m$. Thus 1.7 (2) is proved. $\square$

1.12 Lemma. The diagonal is non-standard.

Proof. Clearly, for each n , $[d]_U \neq [k_n]_U$ since the set $\{i \mid d(i) = n\}$ is a one-element set. This completes the proof of 1.7. $\square$

1.13 Remark. In subsection (e) we shall generalize this construction in two ways: first, we shall start with a possibly non-standard model M instead of N and, second, we shall only assume that M is a model of a fragment of PA .

(c) On Submodels and Cuts

This subsection is somewhat technical; we collect here several facts useful further on. They concern three things; first we formulate very useful principles of overspill and underspill, second we introduce the notion of a k -elementary submodel and prove some lemmas about this notion and third, for $M \subseteq K$ we define $\sup_K(M) = \{a \in K \mid (\exists b \in M) M \models a \leq b\}$ and prove some facts about this notion.

1.14 Definition. Let $M \models I_{\text{open}}$. A set $I \subseteq M$ is a *cut* (notation: $I \subseteq_e M$) if it is closed under successor (in the sense of M) and contains with each a each $b \in M$ that is less than a , i.e.

$$(\forall a \in I)[S_M(a) \in I \ \& \ (\forall b \in M)(M \models b < a \rightarrow b \in I)].$$

A cut $I \subseteq M$ is *proper* if $I \neq M$.

1.15 Observation. Let $M \models I_{\text{open}}$. $M \models I\Sigma_k$ iff no proper cut is $\Sigma_0(\Sigma_k)$ -definable.

Proof. (ii) $\Rightarrow$ (i) is evident; and (i) $\Rightarrow$ (ii) follows from the fact that $I\Sigma_k$ implies $I\Sigma_0(\Sigma_k)$ (see 1.2.4). $\square$

1.16 Corollary. Let $M \models I\Sigma_k$. Let $\varphi(x) \in \Sigma_0(\Sigma_k)$ and let $I \subseteq M$ be a proper cut. Then

- (i) (overspill) if for each $a \in I$, $M \models \varphi(a)$ then for some $b \in M - I$, $M \models (\forall x \leq b)\varphi(x)$;
- (ii) (underspill) if for each $a \in M - I$, $M \models \varphi(a)$ then for some $b \in I$, $M \models (\forall x > b)\varphi(x)$.

Proof. If for each $b \in M - I$, $M \models (\exists x < b)\neg\varphi(x)$ then $I = \{a \in M \mid M \models (\forall x < a)\varphi(x)\}$; thus I is $\Sigma_0(\Sigma_k)$ -definable. Similarly for (ii). $\square$

1.17 Definition. Let $M \subseteq K$. M is a k -elementary substructure of K (notation: $M \prec_k K$) if for each $\varphi(x_1, \dots, x_n) \in \Sigma_k$ and each $a_1, \dots, a_n \in M$, $M \models \varphi(a_1, \dots, a_n)$ iff $K \models \varphi(a_1, \dots, a_n)$. M is an elementary substructure of K if $M \prec_k K$ for each k .

1.18 Remark and Notation. Clearly if $M \subseteq_e K$ (i.e. K is an end-extension of M) then $M \prec_0 K$. See below for a stronger result. We write $M \prec_{e,k} K$ for $M \subseteq_e K$ & $M \prec_k K$.

1.19 Lemma. Let $k \geq 1$, and $M, K \models I_{\text{open}}$, $M \subseteq K$. Then $M \prec_k K$ iff, for each Π_{k-1} -formula $\pi(x, y)$ and each $b \in M$ such that $K \models (\exists x)\pi(x, b)$, there is an $e \in M$ such that $K \models \pi(e, b)$.

Proof. The implication $\Rightarrow$ is obvious; let us prove the converse. Assume the condition and prove $M \prec_i K$ by induction on $i \leq k$. For $i = 0$ use induction on the complexity of the formula in question. $\square$

1.20 Lemma. Let $M, K \models I\Sigma_1$ and $M \subseteq K$. Then $M \prec_0 K$.

Proof. Let $\varphi(x)$ be Σ_0 ; then both φ and $\neg\varphi$ are Σ_1 . By Matiyasevič's theorem (provable in $I\Sigma_1$, cf. I.3.25) there are pure existential formulas $\sigma_1(x)$ and $\sigma_2(x)$ such that $I\Sigma_1 \vdash \varphi(x) \equiv \sigma_1(x)$ and $I\Sigma_1 \vdash \neg\varphi(x) \equiv \sigma_2(x)$. Let $a \in M$ and assume $M \models \varphi(a)$. Then $M \models \sigma_1(a)$, thus $K \models \sigma_1(a)$ and $K \models \varphi(a)$. Similarly for $\neg\varphi$. $\square$

1.21 Remark. (1) Note that $I\Sigma_0(\exp)$ proves Matiyasevič's theorem as well (cf. I.3.26); thus $I\Sigma_1$ may be replaced by $I\Sigma_0(\exp)$ in the previous theorem.

(2) Note that for $k \geq 1$, $B\Sigma_k$ and $P\Sigma_k$ are axiomatized by Π_{k+2} -formulas (see Chap. I, Sect. 2 and subsection (d)) below. Furthermore observe that $M \prec_k K$ implies that M satisfies each Π_{k+1} sentence true in K . Thus if $M \prec_{k+1} K$ and K is a model of $I\Sigma_k$, $B\Sigma_k$, $P\Sigma_k$ respectively, then so is M . (Furthermore, $I\Sigma_0$ is Π_1 -axiomatized; thus $M \prec_0 K \models I\Sigma_0$ implies $M \models I\Sigma_0$.)

1.22 Lemma. Let $M \prec_{e,k} K \models I\Sigma_k$. If $M \neq K$ then $M \models B\Sigma_{k+1}$.

Proof. Clearly, $M \models I\Sigma_0$. Let $\pi(x, y, z) \in \Pi_k$, $a, b \in M$, $M \models (\forall x < a)(\exists y)\pi(x, y, b)$. For each $c \in K - M$, $K \models (\forall x < a)(\exists y < c)\pi(x, y, b)$. By underspill, there is a $c \in M$ such that $K \models (\forall x < a)(\exists y < c)\pi(x, y, b)$; thus $M \models (\exists t)(\forall x < a)(\exists y < t)\pi(x, y, b)$. $\square$

1.23 Definition. Let $M \subseteq K$; put $\sup_K(M) = \{a \in K \mid (\exists b \in M)(K \models a < b)\}$. (The cut in K determined by M .)

1.24 Theorem. Let $k \geq 1$, $M \models B\Sigma_k$ and $M \subseteq K \models I\Sigma_1$. Then $M \prec_{k+1} \sup_K(M)$.

Proof. Put $I = \sup_K(M)$. Clearly $I \prec_0 K$ (since $I \subseteq_e K$), and $M \prec_0 K$ by 1.20. Thus $M \prec_0 I$. We prove $M \prec_1 I$. Let $(\forall x)\varphi(x, y)$ be Π_1 and assume $M \models (\forall x)\varphi(x, a)$. Let $c \in I$, take a $c' \in M$, $I \models c < c'$. Then $M \models (\forall x < c')\varphi(x, a)$, thus $I \models (\forall x < c')\varphi(x, a)$, thus $I \models \varphi(c, a)$. We have proved $I \models (\forall x)\varphi(x, a)$. Now let $k \geq 1$, i.e. $k + 1 > 1$; assume the theorem for k , let $M \models B\Sigma_{k+1}$, $M \subseteq K \models I\Sigma_1$, let Φ be a Π_{k+1} -formula $(\forall x)(\exists y)\pi(x, y, z)$ and let $M \models (\forall x)(\exists y)\pi(x, y, a)$. By induction, we assume $M \prec_k I$. Let c, c' be as above; since $M \models B\Sigma_{k+1}$, there is a $d \in M$ such that $M \models (\forall x < c')\pi(x, (d)_x, a)$. Thus $K \models (\forall x < c')\pi(x, (d)_y, a)$ and $K \models (\exists y)\pi(x, y, a)$. We have proved $K \models (\forall x)(\exists y)\pi(x, y, a)$. $\square$

1.25 Remark. (1) In the next subsection we shall show that the above theorem is the best possible: for each $k > 1$, there are $M \models B\Sigma_k$ and $K \models I\Sigma_1$ such that for $I = \sup_K(M)$ M is not a $(k + 2)$ -elementary substructure of I . (See 1.51)

(2) For PA we get the following Corollary: If $M, K \models PA$, $M \subseteq K$ and $I = \sup_K(M)$, then $M \prec I$ (M is an elementary substructure of I).

(d) Models for the Hierarchy

In this section we shall show that the hierarchy of theories $I\Sigma_k$, $B\Sigma_k$ is proper; we shall also show the exact place of theories $P\Sigma_k$ investigated in Chap. I, Sect. 2.

Recall that $P\Sigma_k$ is $I\Sigma_1$ plus the schema saying that each Σ_k -definable partial function has arbitrarily long finite approximations, i.e. if F is such a function then for each x there is a segment s of length x such that, for each $i \leq x - 1$ and each $u < (s)_i$, u from $\text{dom}(F)$, we have $F(u) < (s)_{i+1}$. Recall the positive results we have:

- 1.26 Theorem.** (1) $PA = \bigcup_n I\Sigma_n = \bigcup_n B\Sigma_n = \bigcup_n P\Sigma_n$
 (2) for $n \geq 0$, $I\Sigma_{n+1}$ implies $B\Sigma_{n+1}$ and $B\Sigma_{n+1}$ implies $I\Sigma_n$;
 (3) for $n \geq 1$, $I\Sigma_{n+1}$ implies $P\Sigma_n$ and $P\Sigma_n$ implies $I\Sigma_n$;

(4) $P\Sigma_0$ is equivalent to $P\Sigma_1$ and $B\Sigma_0$ is equivalent to $B\Sigma_1$.

Recall also the following fact (more or less explicit in Chap. I, Sect. 2, see also the preceding subsection).

1.27 Theorem. (1) $I\Sigma_0$ is a Π_1 -theory.

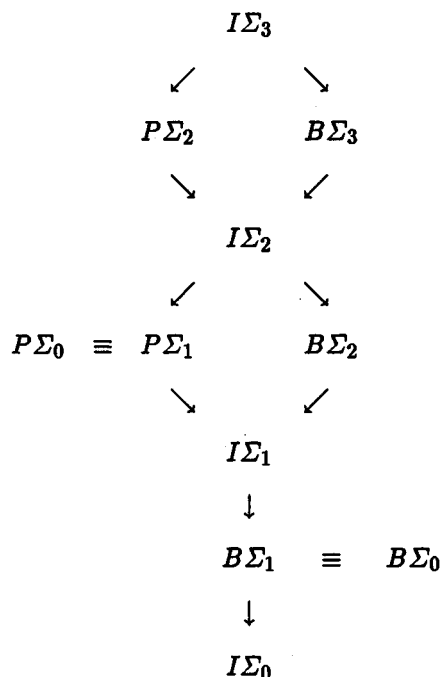
(2) For $n \geq 1$, each of $I\Sigma_n$, $B\Sigma_n$, $P\Sigma_n$ is a Π_{n+2} -theory (i.e. it has a system of axioms consisting only of Π_{n+2} formulas).

Remark. The proof is an easy exercise; note that for $n > 1$, $B\Sigma_n$ is equivalent to $I\Sigma_1$ plus

$$(\dots)(\forall a)[(\forall x < a)(\exists y)\varphi(x, y) \rightarrow (\exists s)(\forall x < a)\varphi(x, (s)_x)]$$

(cf. I.2.24).

1.28. Thus the relative strength of the theories in question may be visualized by the following diagram where $T \rightarrow S$ means: T is at least as strong as S (cf. I.2.26):



We shall show that the diagram remains valid if the arrow is understood as meaning "is strictly stronger than". In more details:

1.29 Theorem. (1) For $n \geq 1$, $I\Sigma_{n+1}$ is strictly stronger than either of $B\Sigma_{n+1}$ and $P\Sigma_n$; both $B\Sigma_{n+1}$ and $P\Sigma_n$ are strictly stronger than $I\Sigma_n$.

Furthermore, $I\Sigma_1$ is strictly stronger than $B\Sigma_1$ and $B\Sigma_1$ is strictly stronger than $I\Sigma_0$.

(2) For $n \geq 1$, $P\Sigma_n$ and $B\Sigma_{n+1}$ are incomparable, i.e. neither of them is a subtheory of the other.

(3) For $n \geq 1$, $P\Sigma_n$ & $B\Sigma_{n+1}$ does not imply $I\Sigma_{n+1}$.

1.30. To prove this theorem we define three constructions of submodels of a given non-standard model; all are given by some definable elements. Note that an element $a \in M$ is Σ_n -definable if there is a Σ_n -formula $\varphi(x)$ such that a is the unique element of M satisfying φ in M . More generally, a is Σ_n -definable from $\mathbf{b}$ in M (where $\mathbf{b}$ is a tuple of elements of M) if there is a Σ_n -formula $\varphi(x, \mathbf{y})$ such that a is the unique $c \in M$ such that $M \models \varphi(c, \mathbf{b})$. If $X \subseteq M$ then a is called Σ_n -definable from X in M if it is Σ_n -definable from a tuple of elements of X in M .

Clearly, each standard element of an $M \models I_{\text{open}}$ (or even $M \models Q$) is Σ_0 -definable ($\bar{n}$ is definable by $x = \bar{n}$). Our starting point is the following.

1.31 Observation. There is a model $M \models PA$ containing non-standard Σ_0 -definable elements.

Proof. By Gödel's first incompleteness theorem, there is a Π_1 -formula $(\forall x)\varphi(x)$ unprovable in PA ; thus $PA + (\exists x)\neg\varphi(x)$ has a model M . Let a be the least element of M satisfying $\neg\varphi$; then a is defined by $\neg\varphi(x) \ \& \ (\forall y < x)\varphi(y)$ which is Σ_0 . $\square$

1.32 Definition. Let $M \models I\Sigma_1$ and $k \geq 0$.

$$\begin{aligned} K^k(M) &= \{a \in M \mid a \text{ is } \Sigma_k\text{-definable in } M\}, \\ I^k(M) &= \sup_M(K^k(M)) = \\ &\{a \in M \mid a \text{ is majorized by a } \Sigma_k\text{-definable element of } M\}. \end{aligned}$$

If $X \subseteq M$ then $K^k(M, X)$, $I^k(M, X)$ are defined as $K^k(M)$, $I^k(M)$ but replacing "definable" by "definable from X ".

$H^k(M) = \bigcup_{n=0} H_n^k(M)$ where $H_0^k(M)$ is $I^k(M)$ and $H_{n+1}^k(M)$ is $I^k(M, H_n^k(M))$.

1.33 Theorem. (1) Let $k > 0$ and $M \models I\Sigma_k$. Then

- (i) $K^k(M) \prec_k M$, thus $K^k(M) \models Th_{\Pi, k+1}(M)$
- (ii) $I^k(M) \prec_{k-1} M$ and $I^k(M) \models Th_{\Pi, k+1}(M)$
- (iii) $H^k(M) \prec_k M$, thus $H^k(M) \models Th_{\Pi, k+1}(M)$.

Evidently $I^k(M) \subseteq_e M$ and $H^k(M) \subseteq_e M$.

- (2) If $N \neq K^k(M)$ (i.e. if $K^k(M)$ is non-standard) then
- (i) $K^k(M)$ is a model of $I\Sigma_{k-1}$ and $\neg B\Sigma_k$; if $k > 1$ then also $K^k(M) \models P\Sigma_{k-1}$.
 - (ii) $I^k(M)$ is a model of $B\Sigma_k$ and $\neg I\Sigma_k$; if $k > 1$ then also $I^k(M) \models \neg P\Sigma_{k-1}$.
 - (iii) $H^k(M)$ is a model of $B\Sigma_{k+1}$ and $\neg P\Sigma_k$; thus $H^k(M) \models \neg I\Sigma_{k+1}$.

Remark. Assertion (1) and the positive parts of assertion (2) relativize, i.e. $K^k(M)$ may be replaced by $K^k(M, X)$ etc.; but trivially, the negative parts do not, e.g. if $X = M$ then $K^k(M, X) = M$ and $M \models I\Sigma_k$ by assumption; thus $M \models B\Sigma_k$.

1.34 Corollary. (1) If $M \models PA$ and $K^\infty(M)$ means the substructure of all definable elements then $K^\infty(M) \prec M$.

(2) Theorem 1.29 follows.

(3) PA is not finitely axiomatizable, since each finite subtheory of PA is a subtheory of some $I\Sigma_k$. (Cf. 3.2.24)

The rest of the subsection contains a proof of Theorem 1.33 and a proof of the fact that Theorem 1.24 in the previous subsection cannot be strengthened.

1.35 Definition (only for this section). A Σ_k -formula $\varphi(x)$ (possibly containing other free variables as parameters) is *special* if $I\Sigma_0 \vdash \varphi(x) \ \& \ \varphi(y) : \rightarrow x = y$, i.e. in each model of $I\Sigma_0$, at most one object satisfies φ .

1.36 Lemma. (1) An element $a \in M$ is Σ_k -definable in $M \models I\Sigma_k$ iff it is defined by a special Σ_k -formula.

(2) More generally, if $X \subseteq M$ is Σ_k -definable and non-empty in $M \models I\Sigma_k$ then there is a special Σ_k -formula defining an element of X .

Proof. Let $\varphi \in \Sigma_k$ define $X \subseteq M$. If $k = 0$ then $\min X$ has a special Σ_0 definition. Assume $k > 0$ and let $\varphi(x)$ be $(\exists y)\psi(x, y)$; let $\chi(x)$ be $(\exists y, z)\psi(x, y) \ \& \ z = \langle x, y \rangle \ \& \ (\forall u < z)(\neg\psi((u)_0, (u)_1))$. Then χ is special and clearly defines an element of X (not necessarily the least element). $\square$

1.37 Lemma. If $M \models I\Sigma_k$ ($k \geq 1$) is non-standard then Σ_k -definable elements are not cofinal (i.e. there is an $a \in M$ such that no $b \in M$, $b >_M a$ is Σ_k -definable). This generalizes to Σ_k -definability from a proper cut.

Proof. Formalize the notion of a special definition in $I\Sigma_1$. In M , let $y = F(x)$ if x is a special Σ_k^* -formula^{*} and y satisfies x (w.r.t. $Sat_{\Sigma, k}$). Clearly, F is a Σ_k -definable partial function. If a is non-standard then, by $S\Sigma_k$, there is a t such that $(\forall x < a)(x \in \text{dom } F : \rightarrow F(x) < t)$. If b is Σ_k -defined in M by a special Σ_k -formula φ then $M \models b = F(\bar{\varphi}) \ \& \ \bar{\varphi} < a$, thus $M \models b < t$. $\square$

1.38 Convention. In the following lemmas 1.39–1.44 we assume $k \geq 1$, $M \models I\Sigma_k$, $X \subseteq M$.

1.39 Lemma. $K^k(M) \prec_k M$.
(The same for $K^k(M, X)$.)

Proof. Use lemma 1.19: let $M \models (\exists y)\pi(a, y)$, $\pi \in \Pi_{k-1}$, $a \in K^k$, a defined by $\theta \in \Sigma_k$; the least y such that $\pi(a, y)$ is defined in M by

$$(\exists u)(\theta(u) \& \pi(u, y) \& (\forall y' < y) \neg \pi(u, y')),$$

the last formula is Σ_k in $I\Sigma_k$ (even in $B\Sigma_{k-1}$) and such a y exists by L_{k-1} . Thus the least y such that $\pi(a, y)$ is in M . $\square$

1.40 Corollary. $K^k(M, X)$ satisfies $I\Sigma_{k-1}$; if $k \geq 2$ then it also satisfies $P\Sigma_{k-1}$.

1.41 Lemma. If $N \neq K^k(M)$ then $K^k(M) \models \neg B\Sigma_k$.

Proof. Assume $K^k \models B\Sigma_k$. In K^k , each element is Σ_k -definable (by any special Σ_k -formula defining this element in M). Thus for $s, t \in K^k - N$, $K^k \models s < t$, we have (after an obvious formalization, see above)

$$K^k \models (\forall x < t)(\exists e < s)(e \text{ is a special } \Sigma_k^\bullet\text{-definition of } x).$$

By $B\Sigma_k$ the last formula is equivalent to a Σ_k -formula and consequently it is true in M . The restriction of the relation “ e is a special Σ_k -definition of x ” to $s \times t$ is coded (by $I\Sigma_k$) and yields a finite set which is a one-one mapping from $(< s)$ onto $(< t)$, which contradicts the pigeon-hole principle for finite sets, provable in $I\Sigma_1$. This completes the proof. $\square$

1.42 Lemma. $I^k(M, X) \prec_{k-1} M$.

Proof. Trivial for $k = 1$ since $I^k(M, X) \subseteq_e M$. Suppose $k \geq 2$, let $\varphi \in \Pi_{k-2}$ and $M \models (\exists x)\varphi(x, a)$ where $a \in I^k$. (Write I^k instead of $I^k(M, X)$.) We find a $d \in K^k$ such that $M \models (\exists x < d)\varphi(x, a)$; and since such an x is necessarily in I^k , the lemma follows by 1.19. Thus let $c \in K^k$ be such that $M \models a < c$; in M , take $d = \max_{y \leq c} \min_x \varphi(x, y)$. In more detail, for $y \leq c$ put $F(y) = \min_x \varphi(x, y)$ if there is such an x . Clearly, F is $\Sigma_0(\Pi_{n-2})$ -defined and therefore coded; let $d = \max(\text{range}(F))$. Then d is defined as follows:

$$\begin{aligned} x = d \equiv & (\exists y \leq c)(\varphi(x, y) \& (\forall y' < y) \neg \varphi(x, y')) \\ & \& (\forall y \leq c)(\forall z)(\varphi(z, y) \rightarrow (\exists z' \leq x)\varphi(z', y)). \end{aligned}$$

Write this as $x = d \equiv \alpha(x, c)$; clearly α is Σ_k (even Π_{k-1}) in M ; if $\gamma(y)$ is a Σ_k -definition of c , d is Σ_k -defined by $(\exists y)(\bar{\gamma}(y) \ \& \ \alpha(x, y))$. $\square$

1.43 Lemma. $I^k(M, X) \models Th_{\Pi, k+1}(M)$, i.e. each Π_{k+2} sentence true in M is true in $I^k(M, X)$.

Proof. Let π be Π_{k-1} and let $M \models (\forall x)(\exists y)\pi(x, y)$. Take an arbitrary $a \in K^k$. Then $M \models (\forall x < a)(\exists y)\pi(x, y)$ and, by $B\Sigma_k$, $M \models (\exists t)(\forall x < a)(\exists y < t)\pi(x, y)$. We show that the least such t_0 is in K . Indeed t_0 is $\Sigma_0(\Pi_{k-1})$ in a , thus Σ_k -definable (without parameters) in M , hence $t_0 \in K$. Thus $I^k \models (\forall x < a)(\exists y < t_0)\pi(x, y)$ (thanks to $I^k \prec_{k-1} M$) and since a was an arbitrary element of K we get $I^k \models (\forall x)(\exists y)\pi(x, y)$. $\square$

1.44 Lemma. (1) $I^k(M, X) \models B\Sigma_k$.

(2) If $k \geq 2$ then $I^k(M, X) \models P\Sigma_{k-1}$.

(3) If X is a proper cut in M or if $X = \emptyset$ and M is non-standard then $I^k(M, X) \models I\Sigma_k$, thus $I^k(M, X) \neq M$.

Proof. (1) follows by Lemma 1.22 for $I^k \neq M$ and is trivial for $I^k = M$. (2) follows by 1.33 since $P\Sigma_{k-1}$ is Π_{k+1} . (3) follows by 1.37. $\square$

1.45 Lemma. If $k \geq 1$, and $M \models B\Sigma_{k+1}$ then (1) $H^k(M) \models B\Sigma_{k+1}$ and (2) $H^k(M) \prec_k M$.

Proof. (1) follows by Lemma 1.22 for $H^k \neq M$ and is trivial for $H^k = M$ (since $M \models B\Sigma_{k+1}$). (2) is clear since $H^k(M)$ is the sum of the increasing chain of models $K^k(H_n^k(M))$ and each such model is $\prec_k M$. $\square$

1.46 Remark. (1) Observe that the assumption $M \models B\Sigma_{k+1}$ may be replaced by $M \models I\Sigma_k$ and $H^k(M) \neq M$.

(2) Recall the notion “a finite increasing sequence is an approximation of F ” (where F is a Σ_k -defined partial function); note that this is a Π_k notion. Thus (1) $I\Sigma_k$ proves that if there is an approximation s of F such that $lh(s) = x$ then there is a least such s ; (2) for each n , $I\Sigma_k$ proves that there is an approximation s of F with $lh(s) = n$. (See I.2.21).

1.47 Definition ($I\Sigma_1$). From here on out, let F be the function associating with each pair (z, x) (where z is a special parametrical Σ_k definition and x is a parameter) the only y defined by z from x if there is such a y , and undefined otherwise.

(2) A finite increasing sequence s is said to *capture* x if x is less than or equal to the last member of s .

1.48 Lemma. Let $M \models I\Sigma_k$ and $a \in M$; $a \in H^k(M)$ iff, for some standard n , the least approximation s of F with $lh(s) = n$ captures a .

Proof. Since $H_k(M) \models I\Sigma_k$, for each standard n there is an $s_n \in H_k(M)$ such that, in $H_k(M)$, s_n is the least approximation of F such that $lh(s_n) = n$. Since $H_k(M) \prec_k M$, s_n has the same property in M ; this proves that if a is captured by such an s_n then $a \in H^k(M)$.

Conversely, let us prove by induction on i that each element of H_i^k is captured by some s_n . This is clear for $i = 0$; assume the assertion for i and consider $(i + 1)$. It suffices to show that an element of $K^k(H_i^k)$, i.e. an element a Σ_k -definable from an element c of H_i^k , is captured. Let $\psi \in \Sigma_k$ and ψ define a from c in M ; observe that ψ , being standard, is in H_i^k and hence $(\bar{\psi}, c) \in H_i^k$. By the induction assumption, let s_n capture $(\bar{\psi}, c)$; then clearly, s_{n+1} captures a . $\square$

1.49 Corollary. Let $M \models I\Sigma_k$ be non-standard and let $H^k(M) \neq N$. Then $H^k(M)$ is not a model of $P\Sigma_k$. Thus $H^k(M)$ is not a model of $I\Sigma_{k+1}$.

Proof. In $H^k(M)$, the function F defined in 1.37 has no approximation of non-standard length. $\square$

1.50 Summary. We check that Theorem 1.33 has been proved. (1) (i) by 1.39, (2) (i) by 1.40 and 1.41; (1) (ii) by 1.42 and 1.43; (2) (ii) by 1.44; (1) (iii) by 1.45, (2) (iii) by 1.45 and 1.49.

*

The last task of the present subsection is to show that the result of 1.24 is optimal.

1.51 Theorem. For each $k \geq 1$ there are $M \subseteq K$ such that $M \models I\Sigma_k$, $K \models I\Sigma_{k+1}$ and M is not a $(k + 2)$ -elementary substructure of $\sup_K(M)$.

Proof. Take a $K \models I\Sigma_{k+1}$ containing non-standard Σ_0 -definable elements, and let $M = K^{k+1}(K)$; thus $M \models I\Sigma_k$ & $\neg B\Sigma_{k+1}$ (cf. 1.33). Let $I = \sup_K(M) = I^{k+1}(K)$. Note that in M , each element is Σ_{k+1} -definable; in other words, for each non-standard t ,

$$M \models (\forall x)(\exists y < t)(y \text{ special}^* \text{ and } \text{Sat}_{\Sigma, k+1}(y, x)).$$

Assuming $M \prec_{k+2} I$, the last formula would be true in I for each non-standard t (since M is cofinal in I), thus in I each element would be Σ_{k+1} -definable. But by 1.33, $I \models B\Sigma_{k+1}$ and the following lemma gives a contradiction. $\square$

1.52 Lemma. If $k \geq 1$ and $M \models B\Sigma_{k+1}$ is non-standard then some elements of M are not Σ_{k+1} -definable. More generally, assume $a \in M$ and $t \in M - N$; then under the above assumption, there are elements $b \in M - N$, such that $M \models b \leq t$ and b is not Σ_{k+1} -definable from a .

Proof. Assume that each $b <_M t$ is Σ_{k+1} -definable from a ; thus $M \models (\forall x < t + 1)(\exists y < t)(y \text{ special}^\bullet \text{ and } \text{Sat}_{\Sigma, k+1}(y, x, a))$.

This can be written as

$$M \models (\forall x < t + 1)(\exists y < t)(\exists u)\pi(x, y, u, a)$$

where $\pi \in \Pi_k$. Using $M \models B\Sigma_{k+1}$ we can bound the quantifier $(\exists u)$; thus for some q ,

$$M \models (\forall x < t + 1)(\exists y < t)(\exists u < q)\pi(x, y, u, a).$$

By $I\Sigma_k$, there is a $d \in M$ coding the relation on $(t + 1) \times t$ defined by $(\exists u < q)\pi(x, y, u, a)$. From d we may construct a d' which is, in M , a finite mapping associating to each $x \leq t$ its least special Σ_{k+1} -definition from a . Since $M \models (d' \text{ is one-one})$ we get a contradiction with the pigeon hole principle for finite functions (provable in $I\Sigma_1$). $\square$

(e) Elementary End Extensions

1.53 Theorem. Let $M \models I\Sigma_0$ (countable!).

- (1) For $k \geq 1$, the following are equivalent:
 - (li) $M \models B\Sigma_{k+1}$
 - (lii) M has a proper $(k + 1)$ -elementary end extension $K \models I\Sigma_0$
- (2) If M has a proper 1-elementary end extension $K \models I\Sigma_0$ then $M \models B\Sigma_2$.
- (3) If $M \models I\Sigma_0$ then $[M \text{ has a proper elementary end extension iff } M \models PA]$.

1.54 Remark. The theorem is proved in this subsection. In (1) observe that (lii) evidently implies that the K in question is a model of $I\Sigma_{k-1}$ (since this theory is Π_{k+1}). We don't know if we may always have $K \models B\Sigma_k$; but if $k \geq 2$ we may have $K \models P\Sigma_{k-1}$.

The reader may show as an exercise that (lii) *cannot* be strengthened to $K \models I\Sigma_k$ (this would imply $M \models I\Sigma_{k+1}$). But comparing this theorem with 1.22, we may ask if 1.22 can be converted, i.e. if (li) implies that M has a proper k -elementary end extension to a model of $I\Sigma_k$. The answer is negative (personal communication by R. Kaye).

1.55 (Proof of (2)). Assume $M \models I\Sigma_0$, $M \prec_{e,1} K \models I\Sigma_0$; we prove $B\Pi_1$. Let $\pi(x, y) \equiv (\forall z)\varphi(x, y, z)$ where φ is Σ_0 and assume $M \models (\forall x < a)(\exists y)\pi(x, y)$ (we disregard parameters). Let $t \in K - M$; then $K \models (\forall x < a)(\exists y < t)\pi(x, y)$

– here we use $M \prec_1 K$. For each $c \in K - M$, $K \models (\forall x < a)(\exists y < t)(\forall z < c)\varphi(x, y, z)$.

Fix a and c ; by underspill, there is a $t \in M$ such that the last formula is true in K . Thus, $M \models (\forall x < a)(\exists y < t)\pi(x, y, z)$. $\square$

1.56 (*Proof of (1ii) $\Rightarrow$ (1i).*) Let $k \geq 1$, $M \models I\Sigma_0$, $M \prec_{e, k+1} K \models I\Sigma_0$. By 1.22 and by induction on k , we know that $M \models B\Sigma_k$ and $K \models I\Sigma_{k-1}$. We prove $M \models B\Pi_k$. Let $\pi(x, y) \equiv (\forall z)\varphi(x, y, z)$ where φ is Σ_{k-1} . Assume $M \models (\forall x < a)(\exists y)\pi(x, y)$. For a $t \in K - M$, we have

$$K \models (\forall x < a)(\exists y < t)(\forall z)\varphi(x, y, z).$$

Thus

$$\begin{aligned} K &\models (\forall x < a)(\forall u)(\exists y < t)(\forall z < u)\varphi(x, y, z), \\ K &\models (\exists t)(\forall x < a)(\forall u)(\exists y < t)(\forall z < u)\varphi(x, y, z); \end{aligned}$$

since $K \models B\Sigma_{k-1}$ the last sentence is Σ_{k+1} and therefore holds in M .

$$M \models (\exists t)(\forall x < a)(\forall u)(\exists y < t)(\forall z < u)\varphi(x, y, z).$$

Since $M \models B\Sigma_k$ we may use $R\Sigma_{k-1}$ (regularity) and get

$$M \models (\exists t)(\forall x < a)(\exists y < t)(Cu)(\forall z < u)\varphi(x, y, z)$$

where (Cu) is “for infinitely many u ” (cf. 1.2.20); thus

$$M \models (\exists t)(\forall x < a)(\exists y < t)(\forall z)\varphi(x, y, z), \quad \square$$

1.57 (*Proof of (1i) $\Rightarrow$ (1ii).*) We generalize and refine the construction of the definable ultrapower described in subsection (b) (1.6–1.12). First, we start with an $M \models B\Sigma_{k+1}$, not just by N . Second, we take algebra D of (parametrically) Δ_k -definable subsets of M and construct an ultrafilter U on D . To be able to prove that in the corresponding model constants form an initial segment (cf. 1.11), U must have the following property: each parametrically Δ_k -definable partition of M into a M -finite number of sets intersects U . In other words, if f is a Δ_k -definable mapping of M into $(< a)$ where $a \in M$ then for some $b <_M a$, $f^{-1}(b) \in U$. Due to countability, we can arrange all such F ’s (satisfying the “if”-part above) into a sequence $f_0, f_1, \dots$ and define: $X_0 = M$; and if X_n has been defined and is unbounded in M , put $f' = f_n \upharpoonright X_n$. Since $M \models B\Sigma_{k+1}$, it satisfies $\omega \rightarrow \omega'_{<\omega}(\Delta_k, \Delta_k)$ and therefore there is a b such that $f'^{-1}(b)$ is unbounded; we put $X_{n+1} = f'^{-1}(b)$. Then we put, for $A \in D$, $A \in U$ iff $(\exists n)(X_n \subseteq A)$.

Let F be the set of all parametrically Δ_k -definable mappings of M into itself. Define $f =_U y$, $[f]_U$ as in 1.6 and let $K = \{[f]_U \mid f \in F\}$. Define $\leq_K$, $+_K$, $*_K$, etc. as in 1.6. We have the following

Claim 1. Łoś's theorem holds for Σ_k -formulas, i.e. for each such formula $\varphi(x_1, \dots, x_n)$ and $f_1, \dots, f_n \in F$, $K \models \varphi([f]_U, \dots, [f_n]_U)$ iff there is an $X \in U$ such that

$$(\forall a \in X)(M \models \varphi(f_1(a), \dots, f_n(a))).$$

To prove this we first prove a subclaim: if φ is a Boolean combination of Σ_{k-1} -formulas and f_i are as above then

$$K \models \varphi([f_1]_U, \dots, [f_n]_U) \text{ iff } \{a \in M \mid M \models \varphi(f_1(a), \dots, f_n(a))\} \in U.$$

The subclaim is proved by induction on the complexity of φ (cf. 1.8–1.9 observing that the sets in question are Δ_k). Then consider $\varphi(x_1, \dots, x_n) \equiv (\exists y)\psi(x, y)$ where ψ is Π_{k-1} .

First assume $K \models (\exists y)\psi([f_1]_U, \dots, y)$; thus there is a g such that $K \models \psi([f_1]_U, \dots, [g]_U)$. By the subclaim, $\{a \in M \mid M \models \varphi(f_1(a), \dots, g(a))\} \in U$. Conversely, let for each $a \in X \in U$, $M \models (\exists y)\psi(f_1(a), \dots, y)$. Then define

$$g(a) = \begin{cases} \min_y M \models \psi(f_1(a), \dots, y) & \text{for } a \in X \\ 0 & \text{for } a \in M - X. \end{cases}$$

Clearly, g is Δ_k -defined and total and $K \models \psi([f_1]_U, \dots, [g]_U)$. This proves Claim 1.

Claim 2. The mapping associating with each $a \in M$ the constant function k_a on M with range $(k_a) = \{a\}$ is a $(k+1)$ -elementary embedding of M onto a proper initial segment of K .

The proof of the fact that this is an embedding of M onto a proper initial segment of K is fully analogous to 1.11–12 (due to our careful choice of U). Furthermore, it is immediate from Claim 1 that for each $\varphi \in \Sigma_k$, $M \models \varphi(a, \dots, b)$ iff $K \models \varphi([k_a]_U, \dots, [k_b]_U)$. Now let $\varphi \in \Sigma_{k+1}$, $\varphi(x_1, \dots, y)$ and let $K \models \varphi([k_a]_U, \dots)$, i.e. for some g , $K \models \psi([k_a]_U, \dots, [g]_U)$. Then for some $c \in M$, $M \models \psi(a, \dots, g(c))$ since otherwise we would get $K \models \neg\psi([k_a]_U, \dots, [g]_U)$ from Claim 1. This proves Claim 2.

Since $k \geq 1$ and after the trivial identification of M with its isomorphic image in K , $M \prec_{k+1} K$, we get $K \models I\Sigma_0$ and (1ii) is proved. $\square$

1.58 (Proof of 1.53 (3).) The implication $\Rightarrow$ follows by 1.55 (1). Conversely, if $M \models PA$ we may modify the construction of the definable K by investigating all definable subsets and construct U such that it intersects each definable partition of M into M -finitely many subsets. Then evidently $M \prec K$ and thus $K \models PA$. $\square$

(f) A Conservation Result

The main aim of this subsection is to prove the following:

1.59 Theorem. For each $k \geq 0$, $B\Sigma_{k+1}$ is a Π_{k+2} -conservative extension of $I\Sigma_k$.

The theorem is due to Paris and Friedman, independently. We shall present two proofs. The first is simpler (due to Kaye; apparently this was the original unpublished Friedman's proof) and the second brings additional information. The first proof relies on the following:

1.60 Lemma. For each $k \geq 0$, if $I\Sigma_k + \psi(a)$ is consistent, where ψ is Π_{k+1} and a is a constant, then there is a model $M \models I\Sigma_k + \psi(a)$ which has a proper k -elementary cut containing a .

Clearly, the lemma implies the theorem, using 1.22: if $I\Sigma_k + (\exists x)\psi(x)$ is consistent then the cut I above is a model of $B\Sigma_{k+1} + \psi(a)$. The second proof uses the following theorem, which is of independent interest.

1.61 Theorem. For each $k \geq 0$, each model M of $I\Sigma_k$ has a cofinal $(k+1)$ -elementary extension K to a model of $B\Sigma_{k+1}$.

This again implies 1.60: if $M \models \psi(a)$ where ψ is as above then $K \models \psi(a)$ since $M \prec_{k+1} K$. We have also the following

1.62 Corollary. For each $k \geq 1$, $(B\Sigma_{k+1} + P\Sigma_k)$ is a Π_{k+2} -conservative extension of $P\Sigma_k$.

Proof. Observe that if M and K are as above and $M \models P\Sigma_k$ then $K \models P\Sigma_k$: if F is a partial function Σ_k -definable in K then its restriction to M (i.e. $F \cap M \times M$) is Σ_k -definable in M ; and if $c \in M$ and $M \models (s \text{ is an approximation of } F \text{ with } lh(s) = c)$ then the same is true in K (since $M \prec_{k+1} K$). This suffices because M is cofinal in K . $\square$

The rest of the subsection elaborates proofs of 1.60 and 1.61.

1.63 (Proof of 1.60). First assume $k = 0$ and let $(I\Sigma_0 + \psi(a))$ be consistent where ψ is Π_1 . Let c be a new constant; then the theory $T = (I\Sigma_0 + \psi(a) + \{a^n < c \mid n \in N\})$ is consistent (by compactness). Let $M \models T$ and let $I = a^N = \{b \in M \mid (\exists n \in N) M \models (b < a^n)\}$. Then I is a proper cut in M , $I \prec_0 M$, $I \models \psi(a)$ and $I \models B\Sigma_1$ by 1.22.

Now assume $k > 0$ and work in $I\Sigma_k$. Say that y majorizes witnesses of Σ_k -formulas beneath z with parameters beneath u (in symbols: $MWitn(y, z, u)$) if

the following holds for each Σ_k formula $x \leq z$ of the form $(\exists v_0)\pi(v_0, \dots, v_t)$ and each t -tuple $s < u$: if there is a w such that $Sat_{\Pi, k-1}(\pi, \langle w \rangle \cap s)(a)$ (a witness for the fact that s satisfies x) then such a w exists beneath y . Define $E(z, u)$ be the minimal y such that $MWitn(y, z, u)$. Observe that E is $\Sigma_0(\Sigma_k)$ and total (due to $S\Sigma_k$).

Now take the following theory T :

$$I\Sigma_k, \quad \psi(a_0), \quad \{a_{n+1} = E(n, a_n) \mid n \in N\}, \quad \{c > a_n \mid n \in N\}.$$

Clearly, T is consistent; let $M \models T$ and $I = \sup_n a_n = \{b \in M \mid (\exists n)(M \models b < a_n)\}$. Then $a \in I$, $I \neq M$, $I \prec_k M$ as desired. $\square$

1.64 Remark. The reader may prove the following fact as an exercise: In the above proof, $I = H^k(M, (\leq a_0))$. The rest of the subsection contains a proof of Theorem 1.61. (The proof differs from the original proof by Paris.)

1.65 (Construction). Recall that each theory T has a conservative extension $S = Sk(T)$ having the following *strong Skolem property*: for each formula $\varphi(y, \mathbf{x})$ of S , there is a function symbol F_φ in S such that $S \models (\exists y)\varphi(\mathbf{x}, y) \rightarrow \varphi(\mathbf{x}, F_\varphi(\mathbf{x}))$. Moreover, each model $M \models T$ has an expansion to a model of S . (see 0.15.)

Let $M \models I\Sigma_k$ and let M^* be the expansion of M to a model of $Sk(I\Sigma_k)$. Let B be the Boolean algebra of all subsets of M definable in the language of $Sk(I\Sigma_k)$ (B is countable). Let U be a non-trivial ultrafilter on B containing all upper segments $(> b) = \{a \in M \mid M \models a > b\}$. Let K^* be the definable ultrapower of M^* given by U and let K be the reduct of K^* to the language of arithmetic. Identifying each $a \in M$ with $\bar{a} = [k_a]$ where $k_a : M \rightarrow \{a\}$ we get $M \subseteq K$. Finally, let $I = \sup_K(M) = \{[f] \in K \mid (\exists a \in M)(K \models [f] \leq \bar{a})\}$. We show that I is the desired extension, i.e. $M \subseteq_{cf} I$, $M \prec_{k+1} I$, $I \models B\Sigma_{k+1}$.

1.66 Fact. (1) $M \prec K$ and M is not cofinal in K .

(2) $M \prec_{k+1} I$.

(3) $I \prec_k K$.

Proof. (1) $M \prec K$ follows from the fact that Łoś's theorem holds for all formulas of S :

$$K^* \models \varphi([f_1], \dots, [f_n]) \text{ iff } \{a \in M \mid M^* \models \varphi(f_1(a), \dots, f_n(a))\} \in U.$$

(The verification is left to the reader as an easy exercise.) In particular, for $a_1, \dots, a_n \in M$,

$$K^* \models \varphi(\bar{a}_1, \dots, \bar{a}_n) \text{ iff } M^* \models \varphi(a_1, \dots, a_n),$$

which means, after the identification of M with the isomorphic substructure of K^* , $M^* \prec K^*$. Consequently, $M \prec K$. M is not cofinal in K since the

diagonal d defined by $d(a) = a$ satisfies $K \models d > \bar{a}$ for each $a \in M$ (by the choice of U).

(2) For $k \geq 1$, the assertion follows by 1.24. For $k = 0$ argue as follows: Let $\pi(x)$ be Π_1 and let $M \models \pi(a)$. Then $K \models \pi(a)$ (by $M \prec K$), hence $I \models \pi(a)$ (since $I \subseteq_e K$).

(3) The claim is trivial for $k = 0$. For $k > 0$ use the function $E(u, z)$ from 1.63 ($E(u, z)$ is the least upper bound for a witness of Σ_k -formulas $\leq u$ with parameters $\leq z$). Let $b \in I$ and $K \models (\exists u)\varphi(u, b)$ where $\varphi \in \Pi_{k-1}$. Let $m \in N$ such that $\varphi < m$ (φ is standard). Furthermore, let $a \in M$ be bigger than b ; in M , let $e = E(m, a)$. Since $M \prec K$ we have $K \models e = E(m, a)$, i.e. $K \models (\exists u < e)\varphi(u, b)$. This means that there is a $c \in I$ such that $K \models \varphi(c, b)$. By 1.19, this implies $I \prec_k K$. $\square$

1.67 Corollary. $I \models B\Sigma_{k+1}$, thus I is the desired extension of M . This follows by 1.22 since $I \neq K$ by 1.66 (1) (M is not cofinal in K). This completes the proof of 1.61.

2. Cuts in Models of Arithmetic with a Top

Cuts in models of Peano arithmetic have played an important role in the development of model theory of arithmetic: Paris and Kirby did pioneering work that continued by the indicator theory and a model-theoretic proof of the independence of the Paris-Harrington principle. Later Paris developed the theory of cuts in models of $I\Sigma_1$; this is the topic of our main interest. Important in Paris's work are cuts in models of arithmetic with top (TA'): they present a powerful device for the study of cuts in models of $I\Sigma_1$, which in turn, are very useful in the study of provably recursive functions. This and the next section are devoted to this.

(a) Arithmetic with a Top and Its Models

2.1. Recall the theory BA' – arithmetic without function symbols, introduced in Chap. I, Sect. 2 (e). Its language L' consists of the constant $\bar{0}$, the predicates $=, \leq, S$ (binary), A, M (ternary) having the usual meaning; axioms assure that $\leq$ is a discrete linear order with the least element $\bar{0}$, S, A, M define partial functions having natural properties of successor, addition and multiplication and, in addition, axioms contain the induction schema for Σ' -formulas (bounded formulas of L').

TA' (arithmetic with top) results by adding the axiom “there is a largest element” and $I\Sigma'_0$ results by adding “successor, addition and multiplication

are total functions". The only reason for the present short subsection is to prove the following

2.2 Theorem. (1) If $M \models I\Sigma'_0$ and $a \in M$ then the substructure $M \upharpoonright a$ of M with the domain $\langle a \rangle$ is a model of TA' .

(2) Conversely, if $B \models TA'$ then there is a model $K \models I\Sigma'_0$ and an $a \in M$ such that B is $M \upharpoonright a$.

2.3 Corollary. $I\Sigma'$ and TA' prove the same Π'_1 -sentences. Thus both theories extend $BA' \Pi'_1$ -conservatively.

2.4 Proof of 2.2 – beginning. (1) is trivial; we prove (2). For one-element B the assertion is evident; so we assume that B has at least two elements. Work with elements of B as with generalized digits and consider finite sequences s of elements of B whose last element is not 0. (The length of s is an element of N .) This is the domain of our K . It is ordered lexicographically; $s \leq t$ iff $lh(s) < lh(t)$ or $[lh(s) = lh(t)$ and if i is minimal such that $(s)_i \neq (t)_i$ then $B \models (s)_i < (t)_i]$. This defines the meaning of S . A and M (addition and multiplication) are defined by algorithms of school mathematics, i.e. $K \models A(s, t, w)$ iff there are sequences q (of carries) and u (sum) of length $\max(lh(s), lh(t)) + 1$ satisfying "bitwise" the obvious conditions for addition (q consists only of zeros and ones and several cases are distinguished, e.g., for each $0 < i \leq \max(lh(s), lh(t))$,

$$\begin{aligned} (q)_{i-1} &= 0 \ \& \ A((s)_i, (t)_i, (u)_i) \ \& \ (q)_i = 0 \ \text{or} \\ (q)_{i-1} &= 0 \ \& \ \neg(\exists z)A((s)_i, (t)_i, z) \\ &\quad \& \ (\forall x)(x \text{ minimal such that } \neg(\exists z)A((s)_i, x, z) \rightarrow A(x, (u)_i, (t)_i)) \\ &\quad \& \ (q)_i = 1 \ \text{or} \\ (q)_i &= 1 \ \& \dots \end{aligned}$$

and w results from u by omitting the last element if it is 0, otherwise $w = u$. Conditions for multiplication are clumsier (use $lh(t)$ auxiliary sequences of length $lh(s) + lh(t)$) but are easy to describe. The proof is completed by proving the following lemmas:

2.5 Lemma. Let B_n be the substructure of K consisting of all sequences of length at most n ($n \geq 0$). Then $B_n \subseteq_e K$ and B_n is definable in B using n -tuples of objects of B to code objects of B_n . (Thus we have a formula with $2n$ free variables defining the successor for n -tuples etc.)

Proof by formalizing the above definitions – for a fixed n . □

2.6 Lemma. For each n , B_n is a model of TA' .

Proof. The verification of properties of $\leq$, S , A , M is tedious but straightforward. To verify the least number principle use the fact that by 2.5, each L' -formula with k free variables speaking about B_n can be equivalently replaced by an L' -formula with kn free variables speaking about B . Then use the lexicographic ordering of n copies of the universe inside B . $\square$

2.7 Lemma. $K \models I\Sigma'$, $B_1 \subseteq_e K$ and B_1 is isomorphic to B .

Proof. Since $K = \bigcup B_n$ is a union of an increasing chain of initial segments, basic properties of S , A , M , $\leq$ are verified easily. The least number principle in K reduces to the least number principle in a suitable B_n . Clearly, K has no greatest element. This completes the proof of Theorem 2.2. $\square$

(b) Cuts

2.8 Definition. (1) *Standard models* of BA' are N and non-empty initial segments of N (finite standard models). All other models of BA' are *non-standard*: by the previous, these are just non-standard models of $I\Sigma'$ and their segments given by non-standard elements.

(2) Let $M \models BA'$ and let $I \subseteq M$. I is a *cut* in M if it is a non-empty initial segment without a largest element in the ordering $\leq_M$. I is *closed under the operations* if, for each $a, b \in I$, there are $c, d \in I$ such that $M \models A(a, b, c)$ and $M \models M(a, b, d)$ (addition and multiplication are total on I).

2.9 Remark. (1) Clearly, a $M \models TA'$ is non-standard iff M has a proper cut. A cut $I \subseteq M$ is closed under operations iff $I \models I\Sigma'$.

(2) Recall the Σ_0 predicate $\exp(x, y)$ defining exponentiation in $I\Sigma_0$ as a partial operation. We shall use the same notation for an equivalent Σ'_1 -predicate.

2.10 Definition. A cut $I \models M$ is *short* if I is closed under operations and there is an $a \in M - I$ such that 2^{2^a} exists in M , i.e.

$$M \models (\exists b, c)(\exp(2, a, b) \ \& \ \exp(2, b, c))$$

2.11 Remark. (1) Observe that if $M \models I\Sigma_1$ (or even if $M \models I\Sigma_0 + (\exp)$) then each proper cut is short.

(2) Recall that in $I\Sigma_0$ we have a Σ_0 predicate $\text{bit}(x, z, y)$ saying “the z -th bit in the binary expansion of x is z ”. Informally, we shall use bit as a symbol for the corresponding partial function.

2.12 Lemma. For each Σ_0 -formula $\varphi(x, y)$ BA' proves the following: For each y and each z such that 2^z exists, there is a $w < 2^z$ such that, for each $x \leq z$

$$\text{bit}(w, x) \equiv \varphi(x, y)$$

Proof. Take a z_0 such that 2^{z_0} exists and prove the assertion by induction for $z \leq z_0$ (observing that it is Σ_0): assume that for z we have the respective $w \leq 2^{z_0}$ and that $z + 1 \leq z_0$ and consider $z + 1$. If $\neg\varphi(z + 1, y)$ then take $w' = w$; otherwise take $w' = w + 2^{z+1}$. $\square$

2.13 Definition. (1) Let I be a short cut in $M \models BA'$, let $A \subseteq I$ and $a \in M - I$. a codes A in M if for each $x \in I$,

$$x \in A \text{ iff } \text{bit}(x, a) = 1.$$

(2) $SS_I(M)$ is the set of all $A \subseteq I$ coded in M .

(3) a codes an $R \subseteq I^k$ iff it codes the set A of all $(a_1, \dots, a_k)$ such that $R(a_1, \dots, a_k)$. ($(\dots)$ is the definable k -tupling function; recall that I is closed under it since $I \models I\Sigma_0$).

2.14 Theorem. Let I be a short cut $M \models BA'$. Then

- (1) $SS_I(M)$ is a Boolean subalgebra of the algebra of all subsets of I .
- (2) If $A \in SS_I(M)$ and B is $\Delta'(A)$ in I then $B \in SS_I(M)$.

Proof. In fact, we only need an a such that $I < a < 2^a \in M$; clearly, each $x \in SS_I(M)$ has a code $< 2^a$.

It follows immediately by 2.12 that $SS_I(M)$ is closed under Boolean operations and that if $X \in SS_I(M)$ (coded by b) and Y is $\Sigma'(X)$ (in I or in M , which is the same) then $Y \in SS_I(M)$. It remains to assume Y to be both $\Sigma'(X)$ and $\Pi'(X)$ (both in I), $x \in SS_I(M)$. Let $y \in Y \equiv I \models (\exists u)P(u, y) \equiv I \models (\forall v)Q(v, y)$ where $P, Q \in SS_I(M)$. Use the "Rosser device": put $\varphi(y) \equiv (\exists u)P(u, y) \wedge (\forall v < u)\neg Q(v, y)$, i.e. $\varphi(y) \equiv (\exists u)[P(u, y) \& (\forall v < u)Q(v, y)]$. Clearly, $y \in Y$ iff $I \models \varphi(y)$ iff $M \models \varphi(y)$. Thus take a c such that, for all $y < a$,

$$M \models \text{bit}(y, c) = 1 \equiv (\exists u \leq a)(P(u, y) \& (\forall v < u)Q(v, y)).$$

This c codes Y . $\square$

2.15 Definition. (1) A short cut $I \in M$ is *semiregular* if I satisfies strong collection with respect to coded functions, i.e. if $F \in SS_I(M)$ is a function and $a \in I$ then $F''(\leq a)$ is not cofinal in I .

(2) I is a model of $I\Sigma_n^*$ (as a cut in M) if it satisfies induction for all formulas that are $\Sigma_n(X)$ with X ranging over elements of $SS_I(M)$. Similarly for $B\Sigma_n^*$.

2.16 Theorem. Let $M \models BA'$ and let $I \subseteq M$ be a short cut. I is semiregular iff $I \models I\Sigma_1^*$.

Proof. The implication $\Leftarrow$ is obvious since $I\Sigma_1(X)$ proves $S\Sigma_1(X)$. (Recall that $I\Sigma_1(X)$ has a new class variable X and induction for all $\Sigma_1(X)$ formulas, cf. I.2.54) For the converse observe that, by 2.14 (2), we see that I (which is clearly a model of $I\Sigma_0^*$) satisfies $S\Pi_0^*$, therefore satisfies $S\Sigma_1^*$ and hence $I\Sigma_1^*$. (See Chap. I, Sect. 2). $\square$

(c) Extendable, Restrained and Ramsey Cuts

2.17 Definition. (1) Make the following definition in $I\Sigma_1$: X is 1-unbounded if it is unbounded, i.e. $(\forall x)(\exists y > x)(y \in X)$.

A set Y of $(n+1)$ -tuples is $(n+1)$ -unbounded if the set

$$\{x_0 \mid \{\langle x_1, \dots, x_n \rangle \mid \langle x_0, x_1, \dots, x_n \rangle \in Y\} \text{ is } n\text{-unbounded}\}$$

is 1-unbounded.

(2) Similarly, let $M \models BA'$, let I be a short cut in M and let $X \subseteq M^{n+1}$ be coded. We call X $(n+1)$ -unbounded in I if the set

$$\{x_0 \mid \{\langle x_1, \dots, x_n \rangle \mid \langle x_0, \dots, x_n \rangle \in Y\} \text{ is } n\text{-unbounded in } I\}$$

is unbounded in I .

Remark. (1) One can prove that if $k+1 = m+n$ and $m, n \geq 1$ then Y is k -unbounded iff

$$\{\langle x_0, \dots, x_{m-1} \rangle \mid \{\langle x_m, \dots, x_k \rangle \mid \langle x_0, \dots, x_k \rangle \in Y\} \text{ is } n\text{-unbounded}\}$$

is m -unbounded.

(2) The formula saying that the set of all $x_1, \dots, x_n$ satisfying $\varphi(x_1, \dots, x_n)$ is n -unbounded is equivalent to $(Cx_1) \dots (Cx_n)\varphi(x_1, \dots, x_n)$.

2.18 Lemma. For $m, n \geq 1$, let $Rams(n, m)$ be the formula $(\forall F \in \Delta_m)((F : V^n \rightarrow a) \rightarrow (\exists i < a)F^{-1}(i) \text{ is } n\text{-unbounded})$. (This makes sense in $I\Sigma_1$.) In each $y, w > 1$, $I\Sigma_1$ proves

$$B\Sigma_{n+m} \equiv Rams(n, m).$$

Proof. Similarly as in II.1.26, show in $I\Sigma_m$ that $Rams(n, m)$ implies $Rams(n-1, m+1)$ observing that if $F : V^{n-1} \rightarrow a$ is Δ_{m+n} , $F = \lim G$, $G : V^n \rightarrow a$, $G \in \Delta_m$, $\{(x, y) \mid G(x, y) = b\}$ is n -unbounded then

$$\begin{aligned} \{y \mid F \mid y) = b\} &= \{y \mid \lim_x G(x, y) = b\} \\ &= \{y \mid \{x \mid G(x, y) = b\} \text{ is unbounded}\} \end{aligned}$$

is $(n-1)$ -unbounded (cf. the remark in 2.17). Then

$$(\forall m)I\Sigma_1 \vdash \text{Rams}(n, m) \rightarrow B\Sigma_{n+m}$$

by induction on n : for $n=1$ see I.1.23 and for $n>1$ II.1.27.

($\Rightarrow$) Assume $B\Sigma_{n+m}$ and let $F \in \Delta_m$, $F: V^n \rightarrow a$, i.e. F acts on all ordered n -tuples of numbers. Restrict F to increasing n -tuples; you get an $F': [V]^n \rightarrow a$ and, by II.1.6, there is an unbounded A homogeneous for F' . Show that $[A]^n$ is n -unbounded; clearly F is constant on $[A]^n$. $\square$

2.19 Definition. (1) Let $M \models BA'$, $k \geq 1$ and let $I \subseteq M$ be a short cut. (1) I is k -Ramsey in M if for each $a \in I$ and coded $F: I^k \rightarrow a$ there is a $b < a$ such that $F^{-1}(b)$ is k -unbounded in I^k .

(2) Let $I \models I\Sigma'$. A sequence $I \subseteq K_0 \subseteq K_1 \dots \subseteq K_k$ is a k -extending chain over I (notation: $K_0 \prec_I K_1 \prec_I \dots \prec_I K_k$) if

- (i) $K_i \models BA'$ for each $i \leq k$,
- (ii) K_i is a Σ' -elementary substructure of K_{i+1} for each $i < k$ (notation $K_i \prec_0 K_{i+1}$),
- (iii) I is a short cut in K_i for each $i \leq k$,
- (iv) for each $i < k$, there is a $b_{i+1} \in K_{i+1}$ such that $I < b_{i+1} < (K_i - I)$.
(Thus, when going from K_i to K_{i+1} we get no new elements below elements of I but we do get new elements between elements of I and elements of $K_i - I$.)

(3) Let I be a short cut in a model $M \models BA'$. I is k -extendable over M if there is a k -extending chain over I whose 0-th element is M ; I is k -restrainable in M if there is a k -extending chain over I whose last element is M .

2.19 Remark. Note that 1-Ramsey cuts are also called *regular*. The notion of a k -extendable and k -restrainable cut is the key notion in Paris's model-theoretic investigation of the strength of instances of combinatorial problems. Observe that if $K_0, \dots, K_n$ is a k -extending chain over I , $a \in K_0 - I$ and 2^{2^a} exists in K_0 then for $d = 2^{2^a}$ $(K_0 \upharpoonright d), \dots, (K_n \upharpoonright d)$ is a k -extending chain over I which is formed by models of TA' with the same top.

2.20 Theorem. Let $I \subseteq M$ be a short 1-Ramsey cut; then I is semiregular.

Proof. Assume I not semiregular and let H be a coded mapping from $(\leq a)$ into I cofinal in I . H may be partial but we may assume it to be one-one (if not, replace it by $H'(b) = \langle b, H(b) \rangle$, clearly H' is coded and cofinal). Now let, for each $x \in I$, $F(x)$ be the pre-image of the maximal $y \leq x$, $y \in \text{range}(H)$. This is $\Delta_1(H)$, thus F is in $SS_I(M)$; $F: I \rightarrow a$ and clearly, each $F^{-1}(b)$ is bounded in I . $\square$

2.21 Theorem. A short cut $I \subseteq M \models BA'$ is k -Ramsey in M iff $I \models B\Sigma_{k+1}^*$ in M (for each $k \geq 1$).

Proof. Since both conditions imply $I \models I\Sigma_1^*$ the theorem follows from 2.18. $\square$

2.22 Theorem. If a short cut $I \subseteq M \models BA'$ is k -extendable over M then it is k -Ramsey.

Proof. By induction on k .

(1) $k = 1$. Let M, K be a 1-extending sequence over I . We prove $I \models B\Sigma_2^*$. Assume $I \models (\forall x < a)(\exists y)(\forall z)\varphi(x, y, z)$ where φ is $\Sigma_0(X)$ for some coded X . Thus

$$(\forall x < a)(\exists y \in I)(\forall z \in I)I \models \varphi(x, y, z).$$

Here, $I \models$ may be replaced by $M \models$ and therefore, by overspill,

$$(\forall x < a)(\exists y \in I)(\exists z_0 \in M - I)M \models (\forall z < z_0)\varphi(x, y, z).$$

Since $M \prec_0 K$, $M \models$ can be replaced by $K \models$. Thus, for $d \in K$ such that $I < d < M - I$, we have

$$(\forall x < a)(\exists y \in I)K \models (\forall z < d)\varphi(x, y, z).$$

Now let e be minimal in K such that

$$(\forall x < a)(\exists y < e)K \models (\forall z < d)\varphi(x, y, z).$$

Then $e \in I$ (since otherwise $e - 1$ could replace e) and thus

$$I \models (\forall x < a)(\exists y < e)(\forall z)\varphi(x, y, z).$$

This shows $I \models B\Sigma_2^*$ and thus I is 1-Ramsey in M .

(2) Now assume $K_0 \prec_I K_1 \prec_I \dots \prec_I K_{k+1}$ and let $F: I^{k+1} \rightarrow q \in I$ be coded by $f < 2^a$. Thus,

$$(\forall z \in I)K_0 \models (\forall x_0 \dots x_k \leq z)(\exists! u < q)(\langle x_0 \dots x_k, u \rangle \in f).$$

By overspill, there is a $d \in K_0$, $I < d_0$, such that

$$K_0 \models (\forall x_0, \dots, x_k \leq d)(\exists! u < q)(\langle x_0, \dots, x_k, u \rangle \in f).$$

Define, in K_1 , h as the function such that, for $\mathbf{x} = x_0, \dots, x_{k-1}$,

$$K_1 \models (\forall x_0, \dots, x_{k-1} \leq d)(h(\mathbf{x}) = f(\mathbf{x}, d)).$$

This h codes a $H : I^k \rightarrow q$ and by the induction assumption applied to $K_1, \dots, K_{k+1}$, there is a $b < q$ such that $H^{-1}(b)$ is k -unbounded in I^k . We show that $F^{-1}(b)$ is $(k+1)$ -unbounded in I^{k+1} . For this it suffices to show

$$\{x \mid H(x) = b\} \subseteq \{x \mid \{y \mid F(x, y) = b\} \text{ is unbounded}\}.$$

Thus assume $x \in I$, $f(x, d) = b$ and let e be an arbitrarily large element of I . Let $p = \min(f(x, y) = b \text{ and } y \geq e)$ (computed in K_1). By $K_0 \prec_I K_1$, this p must be in K_0 and since $K_1 \models p \leq d$ we have $p \in I$. Thus $\{y \mid F(x, y) = b\}$ is unbounded in I . This completes the proof. $\square$

2.23 Definition. Let $M \models BA'$, let $I \subseteq M$ be a short cut, let $c \in I$. I is *regular* (1-Ramsey) up to c in M if for each $a \leq c$ and each coded $F : I \rightarrow a$ there is an $i < a$ such that $F^{-1}(i)$ is unbounded in I .

I is *semiregular* up to c in M if for each function $F \in SS_I(M)$ and each $a \leq c$, the F -image of $(\leq a)$ is not cofinal in I .

Remark. Compare these definitions with 2.19 and 2.13. It is easy to check that Theorem 2.20 generalizes: if I is regular up to c then it is semiregular up to c .

2.24 Theorem. Let $K_0 \prec_I \dots \prec_I K_k$, $a, b \in K_0$, $a < I < b$, $c \in I$, I semiregular up to c in K_k . Then $K_0 \models [a, b] \xrightarrow{\star} (k+2)_c^{k+1}$.

This theorem will be used in subsection (e). First we prove a lemma.

2.25 Lemma. Let $K_0 \prec_I K_1$, $K_i \models TA'$, $a, b \in K_0$, $a < I < b$, $f \in K_0$, $K_0 \models f : [a, b]^{k+1} \rightarrow c$, $c \in I$. Let $C \subseteq I$ be coded in K_0 and o.t.u. in I , i.e. for each $q \in I$ there is a sequence $s \in I$ coding the first q elements of C . Then there is a $B \subseteq C$, B coded in K_1 , o.t.u. in I and prehomogenous for f , i.e. for arguments from $[B]^{k+1}$, the value of f does not depend on the last coordinate.

Proof. Note that $K_1 \models f : [a, b]^{k+1} \rightarrow c$ and $I \models I\Sigma_1$. (See 2.22, 2.21: $K_0 \prec_I K_1$ implies that I is short, by 2.22 I is 1-Ramsey in K_0 and by 2.21 $I \models B\Sigma_2$.) Let $\pi_0 \in K_1$, $I < \pi_0 < K_0 - I$. Let e code C in K_0 . We show that there is a $\pi \in K_1$ such that $I < \pi < K_0 - I$ and $K_1 \models \pi \in e$. Indeed, for each $x \in I$, $K_1 \models (\exists y \in e)(y < \pi_0 \ \& \ y > x)$, thus, by overspill, there is an $x \in K_1 - I$ such that $K_1 \models (\exists y \in e)(y < \pi_0 \ \& \ y > x)$. Let π be such an y .

If $\beta = \langle \beta_0, \dots, \beta_q \rangle \in K_1$ is an increasing sequence of length $> I$ we call β o.t.u. in I if, for each $r \in I$, $\beta_r \in I$. (Then clearly the set of all elements of the sequence belonging to I is o.t.u. in I .)

(1) In K_1 , define an increasing sequence $\beta = \langle \beta_0, \dots, \beta_q \rangle$ to be a *designated sequence* (d.s.) if $q \geq k$, $\beta_0, \dots, \beta_k$ are the first k elements of e bigger than a

and, for $k < p < q$,

$$K_1 \models \beta_{p+1} = \min_{z \in e} (\forall s \in [\beta_0, \dots, \beta_p]^k) (f(s, z) = f(s, \pi)).$$

(2) We want to show that in K_1 , there is a d.s. β of the length $q > I$ which is o.t.u. in I . This will prove the lemma: B will be the set of elements of β belonging to I .

(3) It follows easily that, in K_1 , for each q there is at most one d.s. of the length q .

(4) Since being a d.s. is Σ' , take in K_1 the maximal d.s. $\beta < \pi$; each other d.s. is an initial segment of β . We show that it has the desired properties.

(5) We show that each d.s. β such that $\beta \in I$ has a proper prolongation which is a member of I . Let $\beta = \langle \beta_0, \dots, \beta_p \rangle \in I$; let $\beta_{p+1} = \min_{z \in e} (K_1 \models (\forall s \in [\beta_0, \dots, \beta_p]^k) (f(s, z) = f(s, \pi)))$. We show that β_{p+1} is in I .

For $s \in [\beta_0, \dots, \beta_p]^k$, let $g(s) = f(s, \pi)$; observe that $g \subseteq [\beta_0, \dots, \beta_p]^k \times (< c)$, thus $g \in I$ and hence $g \in K_0$. For each $d \in K_0$, $d > I$, we get

$$K_1 \models (\exists y < d) (\forall s \in [\beta_0, \dots, \beta_p]^k) (f(s, y) = g(s)),$$

and hence $K_0 \models$ the same. Let u be the smallest witness of the last formula in K_0 ; then u is the smallest witness of it in K_1 , therefore $K_1 \models u < \pi$. But then $u \in I$ and thus $\beta \frown \langle u \rangle$ is the desired prolongation of β in I .

(6) Now we show that if $\beta \in K_1$ is a d.s. and $lh(\beta) \in I$ then $\beta \in I$. Indeed, let $\langle \beta_0, \dots, \beta_q \rangle$ be a d.s. in K_1 and let $q \in I$. For $x_1 < \dots < x_k \leq q$ let $h(x_1, \dots, x_k) = f(\beta_{x_1}, \dots, \beta_{x_k}, \pi)$. Then $h : [q]^k \rightarrow c$ in K_1 , thus $h \in I$ (recall again $I \models I\Sigma_1$).

Since I satisfies even Σ_1 -induction with classes coded in K_0 (due to $K_0 \prec_I K_1$) we may define designated sequences inside I using C and $F = f \upharpoonright [I]^k$. Let $DS'(\beta', a, h)$ be a formula such that $I \models DS'(\beta', s, h)$ iff, in I_1 , $\beta' = \langle \beta'_1, \dots, \beta'_k \rangle$ is a sequence, $\beta'_1, \dots, \beta'_k$ are the first k elements of C bigger than a and, for $k < r < p$, β'_{r+1} is the least z such that

$$(\forall s_1 < \dots < s_k < r) (F(\beta'_1, \dots, \beta'_k) = h(s_1, \dots, s_k)).$$

Clearly, if $p \leq q$ and $\beta' = \langle \beta_0, \dots, \beta' \rangle \in I$ then $I \models DS'(\beta', a, h)$ iff $K_1 \models (\beta'$ is a d.s.).

Now it follows by induction in I that for each $p \leq q$ there is a β' such that $DS'(\beta', s, h)$ ((5) giving the induction step); thus the original d.s. $\beta = \langle \beta_0, \dots, \beta_q \rangle$ must be in I .

(7) Consequently, the maximal d.s. β^* cannot be in I , $lh(\beta^*) \notin I$ and each initial segment of β^* of length $q \in I$ is in I ; this means that β is o.t.u. in I . The set B of all β_q such that $q \in I$ is the desired prehomogenous set.

This completes the proof of the lemma. Now we prove Theorem 2.24. Let $K_0 \prec_I \dots \prec_I K_k$, $a < I < b \in K_0$, $c \in I$. By iterating the construction

of the lemma k times we get an $A \subseteq C$, o.t.u. in I and such that, for $s \in [A]^{k+1}$, the value $f(s)$ depends only on the first coordinate of s . Let g be the corresponding unary function and let h be the function associating with each $i < c$ the first $x \geq a$ such that $g(x) = i$. Since I is semiregular up to c , $h''(\leq c)$ is bounded in I ; let $a_0 \in I$ be an upper bound and put $p = ca_0 + 1$ (recall that $I \models I\Sigma_1$). Let s be the sequence of the p first elements of A ; clearly, $s \in I$. Thus there is a $j < c$ such that $g^{-1}(j)$ is a finite set having at least a_0 elements and, by the choice of a_0 , the minimal element of $g^{-1}(j)$ is $\leq a_0$. Thus $g^{-1}(j)$ is relatively large and homogenous for f . $\square$

(d) Satisfaction in Finite Structures with an Application to Models of $I\Sigma_1$

We shall now formalize our notion of standard models of arithmetic with a top. We gain a notion used in various places in the later text. In the present section we use this notion to prove that each non-standard model of $I\Sigma_1$ is isomorphic to a proper cut of itself.

Recall that in Chap. I, Sect. 1–2 we formalized various sorts of partial satisfaction in the universe: for each k we defined in $I\Sigma_1$ formulas $Sat_{\Sigma,k}$ and $Sat_{\Pi,k}$ (satisfaction for Σ_k and Π_k formulas) and showed that in $B\Sigma_m$ ($m \geq 2$) we have a well-behaving satisfaction for $\Sigma_k(\Delta_m)$ -formulas. Moreover, we showed that in $I\Sigma_1$ we have a well-behaving satisfaction for $\Sigma_k(\Sigma_1)$ formulas. All this was satisfaction in the universe. (Using self-reference one can show that there is no full satisfaction in the universe – this is routine, see III.2.3.)

Now we shall focus our attention to finite structures of the form $[0, x]$ endowed with the standard structure given by the language L' (successor, addition, multiplication as predicates). In $I\Sigma_1$ we show that full satisfaction for such structures exists and is Δ_1 :

2.26 Theorem. ($I\Sigma_1$) There is a Δ_1 relation $x \models z[e]$ (ternary) such that for each x , each L' -formula z and each evaluation e of free variables of z by elements of $[0, x]$, Tarski's conditions for satisfaction in $[0, x]$ hold true.

The *proof* is more or less evident: take x as a parameter and define a notion of partial satisfaction in $[0, x]$ for formulas $\leq z$ as a finite set mapping pairs (z', e') , where $z' \leq z$ is a L' -formula and e' an evaluation as above into $[0, 1]$ such that Tarski's conditions are true; by induction on z show that for each z there is a unique partial satisfaction on $[0, x]$ for formulas $\leq z$ and if q_i is such a satisfaction for z_i ($i = 1, 2$) and $z_1 \leq z_2$ then q_1 is a restriction of q_2 . Then $x \models z[e]$ means that the unique partial satisfaction for $[0, x]$ and formulas $\leq z$ gives the value 1 to the pair (z, e) . $\square$

2.27. We now define an important class of formulas generalizing Σ' -formulas. $\Sigma_{1,0}^G$ formulas are open L' formulas in conjunctive normal form; $\Sigma_{1,2k+1}^G$ formulas have the form $(\exists x)\varphi$ where φ is a conjunction of disjunctions of $\Sigma_{1,2k}^G$ formulas; $\Sigma_{1,2k+2}^G$ formulas have the form $(\forall x \leq y)\varphi$ where φ is a conjunction of disjunctions of $\Sigma_{1,2k+1}^G$ formulas.

A formula is Σ_1^G (generalized Σ') if for some k it is $\Sigma_{1,k}^G$. Thus Σ_1^G formulas result from open L' formulas by finite conjunctions and disjunctions, existential quantification and bounded universal quantification.

The whole definition formalizes easily in $I\Sigma_1$.

2.28 Theorem ($I\Sigma_1$). The Σ_1^G formulas persist upwards on structures of the form $[0, x]$, i.e. for each $z \in \Sigma_1^G$, if $x \leq y$ and $x \models z[e]$ then $y \models z[e]$.

The *proof* by induction is easy. □

2.29 Definition. ($I\Sigma_1$) We define satisfaction for Σ_1^G formulas in the universe ($Sat_{\Sigma_1^G}(z, e)$ or $V \models z[e]$) as follows: for each Σ_1^G formula z and each evaluation e of its free variables by some numbers, $V \models z[e]$ if, for some x (containing all elements of the range of e), $x \models z[e]$.

2.30 Lemma ($I\Sigma_1$). The satisfaction just defined fulfils Tarski's conditions.

The *proof* by induction is easy, observing that the formula for which induction is applied is $\Sigma_0(\Sigma_1)$. Some care is necessary for handling finite conjunctions: if z is $\bigwedge\{\varphi_i, i < a\}$ and for each $i < a$ there is an x such that $x \models \varphi_i[e]$ then, by $B\Sigma_1$, there is an x such that for all $i < a$, $x \models \varphi_i[e]$. Similarly for bounded quantifiers. □

Clearly, each Σ' formula is (equivalent to) a Σ_1^G formula. Conversely, we have the following

2.31 Lemma ($I\Sigma_1$). Each Σ_1^G formula is equivalent to a Σ' formula, i.e. Σ_1^G sets coincide with Σ' sets (and therefore with Σ_1 sets).

Proof evident. □

2.32 Lemma ($I\Sigma_1$). For each x there is a b such that for each $\Sigma_{1,x}^G$ sentence z ,

$$V \models z \text{ iff } b \models z,$$

i.e. z is true in the universe iff z is true in $[0, b]$.

Proof. By possibly renaming variables we may restrict ourselves to $\Sigma_{1,x}^G$ sentences z such that all variables occurring in z are among the first x

variables. Furthermore, we may assume some normal form with respect to conjunctions and disjunctions: if $\bigwedge_{i < a} \bigvee_{j < b} \varphi_{ij}$ is a subformula of z then for each i , all φ_{ij} are pairwise distinct and also all the disjunctions $\bigvee_j \varphi_{ij}$ are pairwise distinct. It follows by induction that all $z \in \Sigma_{1,x}^G$ in this normal form form a finite set c . By $S\Sigma_1$, there is a q_0 such that

$$(\forall z \in c)[(\exists q)(q \models z) \rightarrow (\exists q \leq q_0)(q \models z)];$$

thus, for each $z \in c$, $V \models z$ iff $q_0 \models z$. □

2.33 Remark. We can slightly generalize the preceding by expanding L' by finitely many constants for elements $0, 1, \dots, t$ where t is a fixed number. The resulting language will be denoted $L(t)$.

2.34 Definition ($I\Sigma_1$). Let $\mathbf{a}, \mathbf{b}$ be tuples of numbers of the same length and let q be a number. We write

$$(V, \mathbf{a}) \rightarrow_x (q, \mathbf{b})$$

if for each $z \in \Sigma_{1,x}^G$ with the appropriate number of variables,

$$V \models z[\mathbf{a}] \text{ implies } q \models z[\mathbf{b}].$$

2.35 Corollary of 2.32 ($I\Sigma_1$). $(\forall x)(\exists q)(V \rightarrow_x q)$. (Clearly, $V \rightarrow_x q$ means the same as $(V, \emptyset) \rightarrow_x (q, \emptyset)$, i.e. sentences are concerned.)

2.36 Lemma ($I\Sigma_1$). Assume $(V, \mathbf{a}) \rightarrow (q, \mathbf{b})$ and $x > 0$.

(1) If x is even then

$$(\forall c)(\exists d < q)((V, \mathbf{a}, c) \rightarrow_{x-1} (q, \mathbf{b}, d));$$

(2) if x is odd then

$$(\forall d \leq \max(\mathbf{b}))(\exists c \leq \max(\mathbf{a}))((V, \mathbf{a}, c) \rightarrow_{x-1} (q, \mathbf{b}, d)).$$

(This is a prolongation lemma; think of $\mathbf{a}, \mathbf{b}$ as partial isomorphism. The lemma enables us to prolong it.)

Proof. (1) Let c be given and let Φ be the finite set of all $\varphi \in \Sigma_{1,x-1}^G$ formulas in the above-described normal form such that $V \models \varphi(\mathbf{a}, c)$. Then $V \models (\exists v) \bigwedge \Phi(\mathbf{a}, v)$, hence $q \models (\exists v) \bigwedge \Phi(\mathbf{b}, v)$ and thus there is a d such that, for all $\varphi \in \Phi$, $q \models \varphi(\mathbf{b}, d)$.

(2) Let $d \leq b_i$ and assume $(V, \mathbf{a}, c) \rightarrow_{x-1} (q, \mathbf{b}, d)$ to be false for all $c \leq a_i$. Then for each $c \leq a_i$ there is a formula $u \in \Sigma_{1,x-1}^G$ such that

$$V \models u(a, c) \text{ but } q \models \neg u(b, d).$$

The last condition is Σ_1 ; by $S\Sigma_1$ there is a set Φ such that for each $c \leq a_i$, a formula u satisfying the above can be found in Φ . We can further assume that Φ contains only such formulas u . Then $V \models (\forall x \leq a_i)(\bigvee \Phi(a, x))$ but $q \models \neg \bigvee \Phi(b, d)$, i.e. $q \models (\exists x \leq b_i) \neg \bigvee \Phi(b, x)$, which contradicts $(V, a) \rightarrow_x (q, b)$. $\square$

2.37 Remark. Again, we may generalize to $\Sigma_1^G(t)$, i.e. admit constants for numbers up to t .

2.38 Lemma. If $M \models I\Sigma_1$, $e \in M$ is non-standard and $M \models V \rightarrow_e q$ then there is a cut $I \subseteq M$ not containing q and isomorphic to M .

Proof. By the zig-zag method: let $M = \{a_0, a_1, \dots\}$ be an enumeration of M . Let $c_0 \in M$ be non-standard and let d_0 be such that $d_0 \leq_M q$ and $M \models (V, c_0) \rightarrow_{e-1} (q, d_0)$. Having constructed $c_0, \dots, c_k$ and $d_0, \dots, d_k$ such that

$$M \models (V, c_0, \dots, c_k) \rightarrow_{e-k-1} (q, d_0, \dots),$$

continue as follows: if k is even then take for d_{k+1} the first $a_i \in M$ less than q and not among $d_0, \dots, d_k$; by 2.36 take for c_{k+1} the first a_j such that

$$M \models (V, a_0, \dots, c_{k+1}) \rightarrow_{e-k-2} (V, d_0, \dots, d_{k+1}).$$

(Observe that d_{k+1} is distinct from $c_0, \dots, c_k$ because the corresponding open formula that says it must be true due to the arrow above.)

If k is odd, take for c_{k+1} the first $a_i \in M$ not among $c_0, \dots, c_k$ and find similarly d_{k+1} less than q using 2.36.

If k varies over standard numbers we construct two sequences $(c_0, \dots)$, $(d_0, \dots)$; clearly the mapping $f(c_i) = d_i$ is one-one and maps M onto a cut $I \subseteq M$ not containing q . And f is an isomorphism since for each k , $M \models (V, c_0, \dots, c_k) \rightarrow_0 (q, d_0, \dots, d_k)$ since $e - k - 1$ is non-standard; thus atomic formulas are preserved by f . $\square$

2.39 Remark. Again, we may generalize keeping elements less than a given t fixed. Thus by 2.35 and 2.38 we have proved the following

2.40 Theorem. If $M \models I\Sigma_1$ is non-standard (countable) and $t \in M$ then there is a proper cut $I \subseteq M$ containing t and an isomorphism $f : M \rightarrow I$ identical on $(\leq t)_M$.

(Remark: in fact, we have used $L\Sigma'$ and $I\Sigma'$; but this is immaterial.)

3. *Provably Recursive Functions and the Method of Indicators*

In this section we are going to investigate means necessary to prove that a function is provably recursive. We shall see that the amount of induction necessary to prove that a function is total is related to its growth. We shall give two characterizations of $I\Sigma_k$ -provably recursive functions and two characterizations of PA -provably recursive functions: one using largeness ($\rightarrow^*$) and one using the Schwichtenberg-Wainer hierarchy. As a particular case, we show that $I\Sigma_1$ -provably recursive functions are just primitive recursive functions (Takeuti-Minc). To achieve this, we shall use the method of indicators developed by Paris and Kirby; this method has various other uses but ours is typical. We obtain various independence results (and some few other results) as corollaries. The section is structured as follows: (a) We formulate basic definitions concerning provably recursive functions and present the main result on them. (b) We define indicators and prove the main theorem on them. We also elaborate a technical notion of a Paris sequence (more general than that of Chap. II, Sect. 2) as a means of constructing restrainable cuts. In (c) and (d) we investigate two kinds of indicators, one based on $*$ -largeness and the other on α -largeness. This will give the major results of (a). (e) contains two other corollaries: provability of $(W)_k \rightarrow (PH)_k$ in $I\Sigma_1$ and the fact that each non-standard model of $I\Sigma_1$ has a non-standard cut satisfying PA .

(a) Provably Recursive Functions, Envelopes

3.1 Definition. Let $T \supseteq I\Sigma_1$ and let $\varphi(x, y)$ be a formula; φ is said to *define* a (partial) function in T if

$$T \models (\forall x, y_1, y_2)(\varphi(x, y_1) \& \varphi(x, y_2) \rightarrow y_1 = y_2);$$

φ defines a *total function* in T if, in addition,

$$T \models (\forall x)(\exists y)\varphi(x, y).$$

We also say that the function defined by φ is *T -provably total* (or just that φ is *T -provably total*). (Cf. I.1.51).

3.2 Remark. (1) If φ defines a partial/total function in T then it defines such a function in each model of T ; in particular, if $N \models T$, φ defines a function in N . We shall focus our attention on Σ_1 formulas defining functions in N (recall that we call a function partial recursive iff it has a Σ_1 -definition in N). Clearly, a function may have different definitions and their equivalence may be unprovable in a theory T . Thus, even if we shall freely speak on T -provably

total functions, it should be always clear that we investigate a particular Σ_1 -definition of the function in question. Recall that a Σ_1 -function provably total in T is Δ_1 in T and is called *T-provably recursive*.

(2) Definition 3.1 evidently generalizes to functions with several arguments.

3.3 Definition. Let $T \supseteq I\Sigma_1$. A two-argument function $F(-, -)$ Σ_1 -defined in T is an *envelope* for T -provably recursive functions if

- (1) For each n , the function F_n defined by $F_n(x) = F(n, x)$ is T -provably total and
- (2) For each T -provably recursive function H there is an n such that $T \vdash (\forall x)(H(x) < F_n(x))$; thus H is T -provably majorized by F_n .

Remark. To avoid any misunderstanding, let us stress again that in fact the definition concerns a particular Σ_1 -definition $\varphi(u, x, y)$ of F and the corresponding definitions $\varphi(\bar{n}, x, y)$ of F_n .

On the other hand, if $N \models T$ then φ defines a system $\{F_n \mid n\}$ of total recursive functions such that whenever H is a T -provably recursive function then H is majorized by one of the F_n (and more than that: the last fact is T -provable).

3.4 Lemma. If $F = \{F_n \mid n\}$ is an envelope for T -provably recursive functions then a Σ_1 -definable function H is T -provably total iff, for some n , H is T -provably primitive recursive in F_n (i.e a definition T -provably equivalent to the definition of H may be obtained from the definition of H and definitions of basic primitive recursive functions by composition and primitive recursion).

Proof. $\Leftarrow$ is clear since we assume $T \supseteq I\Sigma_1$ and therefore T -provably recursive functions are closed under primitive recursion. To prove $\Rightarrow$ let $(\exists z)\alpha(x, y, z)$ be a Σ_1 definition of H (α being Σ_0); define $H^*(x) = u$ iff u is a pair $\langle y, z \rangle$ such that $\alpha(x, y, z) \& (\forall z' < z) \neg \alpha(x, y, z')$. Evidently, H^* is a function Σ_0 -defined in T and is T -provably total, hence, for some n , T -provably majorized by F_n . Thus

$$T \models y = H(x) \equiv (\exists z < F_n(x)) \alpha(x, y, z).$$

This shows that H is T -provably primitive recursive in F_n since Σ_0 relations are and bounded minimization (T -provably) preserves (relative) primitive recursiveness. $\square$

Now we formulate our main theorems.

3.5 Theorem. For each $k, n \geq 1$ define (in $I\Sigma_1$)

- (a) $F_{k,n}(x) = \min_y ([x, y] \xrightarrow{*} (k+2)_n^{k+1})$;
 $F_n(x) = \min_y ([x, y] \xrightarrow{*} (n+2)_{n+1}^{n+1})$;

- (b) $G_{k,n}(x) = \min_y([x, y])$ is ω_k^n -large)
 (thus $G_{k,n}(x) = f_{\omega_{k-1}^n}(x+1)$ where f_α is the α -th function of the Schwichtenberg-Wainer hierarchy)
 $G_n(x) = \min_y([x, y])$ is ω_n -large)
 (thus $G_{n+1}(x) = f_{\omega_n}(x+1)$).

Then

- (1) for each k , $\{F_{k,n} \mid n\}$ is an envelope for $I\Sigma_k$ -provably recursive functions.
- (2) for each k , $\{G_{k,n} \mid n\}$ is another such envelope;
- (3) both $\{F_n \mid n\}$ and $\{G_n \mid n\}$ are envelopes for (PA) -provably recursive functions.

3.7 Corollary. $I\Sigma_1$ -provably recursive functions are exactly all primitive recursive functions. (In more detail, φ defines an $I\Sigma_1$ -provably recursive function iff there is a ψ resulting from the definitions of basic primitive recursive functions by finitely many applications of the rule of composition and primitive recursion and such that $I\Sigma_1 \vdash \varphi \equiv \psi$.)

Proof. Each primitive recursive (definition of a) function is $I\Sigma_1$ -provably recursive. On the other hand, it is easy to show for each n in $I\Sigma_1$ that $f_n(x)$ is primitive recursive. Thus the result follows by 3.5 (b) and 3.4. $\square$

3.8 Theorem. (1) For each $k \geq 1$, the following formulas are unprovable in $I\Sigma_k$:

- (a) $(\forall x)(\forall z)(\exists y)([x, y] \rightarrow_{*} (k+2)_z^{k+1})$,
- (b) $(\forall \alpha \preccurlyeq \omega_k)(f_\alpha \text{ is total})$.

(2) The following formulas are unprovable in PA :

- (a) $(\forall x)(\forall z)(\exists y)([x, y] \rightarrow_{*} (z+2)_{z+1}^{z+1})$,
- (b) $(\forall \alpha < \varepsilon_0)(f_\alpha \text{ is total})$.

We postpone the proofs to subsections (c) and (d).

(b) Indicators and Paris Sequences

3.9 Definition. Let $M \models I\Sigma_1$ be countable and non-standard; let χ be a set of cuts in M . A Σ_1 -formula $\varphi(x, y, z)$ is an *indicator* for χ in M if

- (i) φ defines a total two-argument function Y in M ,
- (ii) for each $a, b \in M$ such that $a < b$, $Y(a, b)$ is non-standard iff

$$(\exists I \in \chi)(a \in I \& b \notin I),$$

- (iii) for $a, b, c, d \in M$ such that $c \leq a \leq b \leq d$ we have

$$Y(a, b) \leq Y(c, d).$$

We shall be mostly interested in indicators for cuts $I \models T_2$ in a model $M \models T_1$. If φ is an indicator for such cuts in each (countable) model of T_1 we simply say that φ is an indicator for models of T_2 in T_1 .

The main theorem on indicators follows.

3.10 Theorem. Let $T \supseteq I\Sigma_1$ be a theory (in the language of arithmetic). Let φ be an indicator for T in T and write $Y(x, y) = z$ instead of $\varphi(x, y, z)$. Then

- (i) $T \not\models (\forall x, z)(\exists y)(Y(x, y) \geq z)$;
- (ii) $(\forall n)T \vdash (\forall x)(\exists y)(Y(x, y) \geq n)$;
- (iii) If $N \models T$ then the sentence $(\forall x, z)(\exists y)(Y(x, y) \geq z)$ is neither provable nor refutable in T ;
- (iv) The functions $g_n(x) = \min_y(Y(x, y) \geq n)$ form an envelope for T -provably recursive functions.

Proof. (i) Let $M \models T$ be non-standard, let $N < c < a$. If there is no b in M such that $M \models Y(a, b) > c$ we are done; otherwise let b be minimal with this property. Then there is a cut $a < I < b$ in M such that $I \models T$ and in T there is no b' such that $I \models Y(a, b') > c$ (since otherwise we would have $M \models Y(a, b') > c$, a contradiction).

(ii) Since in each non-standard countable model M of $I\Sigma_1$ proper cuts isomorphic to M are cofinal, given $a \in M$ we find an $a \in I \subseteq_e M$ with $I \models T$; for each $I \leq b \in M$, $Y(a, b)$ must be non-standard, thus $M \models Y(a, b) > n$. We showed $M \models (\forall x)(\exists y)(Y(a, b) > \bar{n})$ and M was arbitrary; thus $T \vdash (\forall x)(\exists y)(Y(a, b) > \bar{n})$.

(iii) is evident.

(iv) Let f be T -provably recursive and assume $T + \{(\exists x)(f(x) \geq g_n(x)) \mid n\}$ to be consistent. Then $T + \{(\exists x)(Y(x, f(x)) \geq \bar{n}) \mid n\}$ is consistent; let M be a model of the latter theory. By overspill, M contains an element a such that the value $Y(a, f(a))$ is non-standard. Consequently, there is a cut $a < I < f(a)$ satisfying T ; in I , f is not total since $f(a)$ is undefined. This contradicts to f being T -provably total. $\square$

3.11 Discussion. We shall be mainly interested in indicators for $I\Sigma_k$ (or for $B\Sigma_{k+1}$). This is why we are interested in a method allowing us to conclude, under some conditions, that between two elements $a < b$ of a model M there is a cut which is k -restrainable in $[0, b]$ (such a cut necessarily satisfies $B\Sigma_{k+1}$, see 2.21–22). If I is such a cut then there are $I \subseteq_e B_k \prec_I \cdots \prec_I B_0 = [0, b]$; for each $i < k$, there is an $h_i \in B_i$ such that $I < h_i < B_{i+1} - I$. Paris exhibited two similar constructions of such a chain and we shall try to isolate their common structure. Both constructions then become particular cases. The idea is to find in $[a, b]$ a coded increasing sequence $\{g_i \mid i < \nu\}$ of non-standard length and another increasing sequence $\langle h_0, \dots, h_{k-1} \rangle$ of elements

bigger than the g_i 's and less than b such that if we put $I = \bigcup_{n \in N} [0, g_n]$ and let B_j be the set of all elements of $[0, b]$ "definable" from $I \cup \{h_j, \dots, h_{k-1}\}$ (w.r.t. an appropriate notion of definability) then $B_{j+1} \prec B_j$, $I \subseteq_e B_j$ and $h_j \in B_j$ for each j and we have to assure that $h_j \notin B_{j+1}$. In more details, we shall have a Δ_1 operation $df(b, i, g, par) \subseteq [0, b]$ and for each j , $B_j = \bigcup_{n \in N} df(b, n, g_n, par_j)$ where $par_j = \langle h_j, \dots, h_{k-1} \rangle$; we have to guarantee $\{u \in B_{j+1} \mid u \notin I \text{ \& } u < h_j\}$ to be empty. To this end it suffices to have, for each $i < \nu$, $[g_{i+1}, h_j] \cap df(b, i, g_i, par_{j+1}) = \emptyset$, as one easily sees. It will turn out that it is enough to have

$$[(g_{i+1}, h_j)] \cap df(b, i, g_i, par_{j+1}) = \emptyset.$$

This leads us to the following definition.

3.12 Definition ($I\Sigma_1$). Let df be a total Σ_1 function such that, for each x , $df(x, i, g, par) \subseteq [0, x]$. Let $b > 0$, $u > 0$ and $h = \langle h_0, \dots, h_{u-1} \rangle$ be a finite increasing sequence of elements $\leq b$. A finite increasing sequence $g = \langle g_0, \dots, g_{\nu-1} \rangle$ of elements $\leq b$ is a *Paris sequence* for h in b w.r.t. df if

- (1) each g_i is less than each h_j ,
- (2) for each $i < \nu - 1$ and each $j < u$,

$$[(g_{i+1}, h_j)] \cap df(b, i, g_i, par_{j+1}) = \emptyset$$

where $par_j = \langle h_j, \dots, h_{u-1} \rangle$.

The number u is called the *dimension* of g (cf. II.2.24 and (c) below).

3.13 Definition. Let $M \models I\Sigma_1$ be non-standard and let, in M , df be a total Σ_1 function such that

$$M \models (\forall x, i, g, par)(df(x, i, g, par) \subseteq [0, x]).$$

Let $b \in M - N$ and let $g, h \in M$ be such that g is a Paris sequence for h in b , $lh(g) = \nu$, $lh(h) = u$. Let $J \subseteq_e M$, $J < \nu$, and let I be the cut $\sup\{g_i \mid i \in J\} = \{a \in M \mid (\exists i \in J)(a < g_i)\}$. Finally, put

$$B_j = \bigcup_{i \in J} df(b, i, g_i, par_j)$$

for $j = 0, \dots, u-1$ (par_j is as above). The operation df is *suitable* for b, g, h, I if

- (1) $j < j' < u$ implies $B_j \supseteq B_{j'}$ and
- (2) each B_j is closed under L' -definability from $B_j \cup I$, i.e. if x is L' -definable in $[0, b]$ by a standard formula from some elements of $B_j \cup I$ then $x \in B_j$.
(Consequently, $I \subseteq_e B_j$.)

3.14 Lemma. Let M, b, g, h, I be as above and let df be suitable for b, g, h, I in M . Then for each standard k such that $M \models k \leq lh(h)$, I is k -restrainable in $[0, b]$.

Proof. The proof is easy since almost everything is in the definitions. Let B_j ($j = 0, \dots$) be as above. Then, evidently, $B_j \supseteq B_{j+1}$, $h_j \in B_j$, and by assumption, $B_j \prec [0, b]$. Thus each $B_{j+1} \prec B_j$. The condition (2) in 3.12 gives, for each $i \in N$ and each $j < k$,

$$[(g_{i+1}, h_j)] \cap df(b, i, g_i, par_{j+1}) = \emptyset;$$

thus there is no x such that $x \in B_{j+1} \& x \notin I$ and $x < h_j$ (x would be in a $DEF(h, i, g_i, par_j)$). Moreover, h_j itself is not in B_{j+1} since if it were, h_{j-1} would also be in B_{j+1} . Thus $I < h_j < B_{j+1} - I$. $\square$

3.15 Corollary. Under the above notation, if u is standard and $u = k$, then $I \models B\Sigma_{k+1}$; if $u > N$ then $I \models PA$.

(c) Paris Sequences of the First Kind

These are the Paris sequences in the sense of II.2.24.

We slightly reformulate the definition and reprove a theorem from there (Lemma 3.17 below). The main result is Theorem 3.20 exhibiting an indicator for models of $B\Sigma_{k+1}$ and Theorem 3.23 (an indicator for models of PA).

3.16 Definition ($I\Sigma_1$). Let $i, g < b$ and let par be an increasing sequence of elements of $[0, b]$. $c \in df_1(b, i, g, par)$ if $c \leq b$ and there is a formula $\varphi(x, y) \in L'$, $\varphi < i$, such that for some parameters e from $[0, g] \cup par$ evaluating y , c is either the minimal or the maximal element of $[0, b]$ satisfying $\varphi(x, e)$ in $[0, b]$. (Clearly, df_1 is Δ_1 defined.)

3.17 Lemma ($I\Sigma_1$, cf. II.2.26). Let $u > 0$ (dimension), let $\nu \geq u$ be such that $2^\nu > \nu^4$, let $c \geq 2$ and let b be such that $[0, b] \xrightarrow{**} (u+2)_c^{u+1}$ (see II.2.7; each $f : [0, b]^{u+1} \rightarrow c$ has a homogenous set Y such that $c \leq \min Y \leq 2^{\min Y} \leq \text{card}(Y)$). Then in $[0, b]$ there are g of length ν and h of length u such that g is a Paris sequence for h w.r.t. df_1 .

Proof (cf. II.2.26). We are going to define a function $F : [0, b]^{u+1} \rightarrow c$ such that if Y is the homogenous set as above and $\beta = \langle \beta_0, \dots, \beta_u \rangle \in [Y]^{u+1}$ then for $h = \langle \beta_1, \dots, \beta_u \rangle$ there is a Paris sequence g of length ν . For each

$\beta \in [0, b]^{u+1}$ define a sequence $(g_i \mid i < \nu)$ (not depending on β_0) as follows:

$$\begin{aligned} \text{par}_j^\beta &= \langle \beta_{j+1}, \dots, \beta_u \rangle, \\ g_0 &= 0, \\ g_{i+1} &= \max[df_1(b, i, g_i, \text{par}_1^\beta) \cap [0, \beta_1]] + 1. \end{aligned}$$

We want to have

$$[\beta_j, \beta_{j+1}] \cap df_1(b, i, g_i, \text{par}_{j+1}) = \emptyset$$

for $j = 0, \dots, u-1$.

Claim 1. The last condition for $j = 0$ guarantees $g_{i+1} \leq \beta_0$ and for $j = 1, \dots, u-1$ guarantees that g is a Paris sequence for $\langle \beta_1, \dots, \beta_u \rangle$. (Cf. II.2.26).

Define $F(\beta)$ as follows: $F(\beta) = (j, i)$ if (j, i) is lexicographically the smallest element of $u \times \nu$ such that

$$[\beta_j, \beta_{j+1}] \cap df_1(b, i, g_i, \text{par}_j) = \emptyset \text{ if there is such } (j, i);$$

Otherwise $F(\beta) = (u, 0)$. Clearly, F can be understood as a mapping into c . Let $Y = \{y_i \mid i < e\}$ be homogenous as above.

Claim 2. The common value of F on $[Y]^{u+1}$ is $(u, 0)$. We prove the claim by contradiction. Assume $F(\beta) = (j, i)$ for a $\beta \in [Y]^{u+1}$ and $j < u$. Let $(g_i \mid i < \nu)$ be the sequence corresponding to β .

(1) Observe that $[\beta_j, \beta_{j+1}] \cap df_1(b, i, g_1, \text{par}_j) \neq \emptyset$.

(2) For our i , $g_i \leq \beta_0$. This is clear for $i = 0$. For $i > 0$ observe that for $\bar{i} < i$ and $\bar{j} < j$ we have $[\beta_{\bar{j}}, \beta_{\bar{j}+1}] \cap df_1(b, \bar{i}, g_{\bar{i}}, \text{par}_{\bar{j}}) = \emptyset$; in particular, take $\bar{i} = i-1$ and $\bar{j} = 0$ and see Claim 1 above.

(3) Since the sequence $(a_i \mid i < \nu)$ does not depend on β_0 , we may conclude that $g_i < y_0$ (y_0 is the smallest member of Y).

(4) Therefore (1) implies that $[\beta_j, \beta_{j+1}] \cap df_1(b, i, y_0, \text{par}_j) \neq \emptyset$. Since $\beta \in [Y]^u$ was arbitrary we have $[y_t, y_{t+1}] \cap df_1(b, i, y_0, \text{par}_j) \neq \emptyset$ for each t such that $j < t < e-u+j-1$. Among those intervals $[y_t, y_{t+1}]$ there are $(|Y|-u)/2$ disjoint intervals, each intersected by an element of $df_1(b, i, y_0, \text{par}_j)$. Thus to get a contradiction, it is sufficient to show the following.

(5) $|df_1(b, i, y_0, \text{par}_j)| < (|Y|-u)/2$. To obtain this, observe that there are at most $(i+1)$ formulas $\leq i$ and each may take parameters from a set of cardinality $y_0 + u + 1$; each pair (formula-parameters) can give at most two elements to $df_1(\dots)$. Thus $|df_1(\dots)| \leq 2(i+1)(y_0 + u + 1)^{(i+1)} \leq 2\nu(y_0 + u + 1)^\nu$. On the other hand, $(|Y|-u)/2 \geq (2^{y_0} - u)/2 \geq 2^{y_0} - 2$ (since $y_0 \geq c \geq u$). Thus we have to prove

$$2^{y_0-2} > 2\nu(y_0 + u + 1)^\nu.$$

Put $x = y_0 + u + 1$; it suffices to prove $2^{x-u-4} > x^\nu$ or (stronger) $2^{x-u-4} \geq x^{2^\nu}$. It is easy to show that this holds e.g. for $x = 2^\nu$ and all bigger x (using $2^\nu \geq \nu^4$). But $y_0 + u + 1 \geq y_0 \geq c \geq 2^\nu$. This completes the proof of Claim 2.

Thus $F(\beta) = (u, 0)$ for all $\beta \in [Y]^{u+1}$. This implies, as in (2) above, that $g_i \leq \beta_0$ for all $i \leq \nu$; and evidently, g satisfies the disjointness conditions from the definition of a Paris sequence for $(\beta_1, \dots, \beta_u)$. This completes the proof. $\square$

3.18 Remark. If we assume in the previous lemma that $c \geq 2^{2^\nu}$ then we may find a Y as above and growing exponentially, i.e. $y_{i+i} \geq 2^{y_i}$. (This follows from II.2.13).

3.19 Lemma. Under the notation of 3.13, if df is df_1 (as defined in 3.16) and $J = N$ then df is suitable.

Proof. The condition 3.13 (1) is satisfied, since clearly $u \subseteq v$ implies $df_1(x, i, g_i, u) \subseteq df_1(x, i, g_i, v)$. Also $I \subseteq_e B_j$ is clear. And if $p \in N$ and $a_1, \dots, a_p \in B_j$, and a is defined from $a_1, \dots, a_p$ by a standard formula φ then one easily writes down a standard formula ψ witnessing, for some standard number k , that $a \in df_1(b, k, g_k, par_j)$. This shows closedness of B_j under Skolem functions. $\square$

3.20 Theorem. For each $k \geq 1$, the function

$$Y_k(a, b) = \max_c([a, b] \rightarrow_* (k+2)_c^{k+1})$$

(or = 0 if such a does not exist)

is an indicator both for models of $I\Sigma_k$ and for models of $B\Sigma_{k+1}$ in $I\Sigma_1$.

Proof. Let $M \models I\Sigma_1$ and let $a < b \in M$.

(a) First assume that there is a cut $a < I < b$ such that $I \models I\Sigma_k$. By II.1.9, for each n , $I \models (\exists y)([a, y] \rightarrow_* (k+2)_n^{k+1})$, thus for each n , $M \models (\exists y < b)([a, y] \rightarrow_* (k+2)_n^{k+1})$ and by overspill we get a non-standard $c < b$ such that $M \models [a, y] \rightarrow_* (k+2)_c^{k+1}$. Thus in M , $Y(a, b)$ is non-standard.

(b) Now assume $Y(a, b) > N$ in M . By II.2.7 (or, better: by an analysis of the proof of II.2.7) there is a non-standard c such that $M \models 2^{2^\nu} \leq c$ and $[0, b] \rightarrow_{**} (k+2)_c^{k+1}$. Apply 3.17 and 3.18: we get g, h such that $h = (h_0, \dots, h_{k-1})$ and g is a Paris sequence of length ν for h w.r.t. df_1 in b ; h is a k -tuple of elements of a set of indiscernibles of exponential growth and non-standard cardinality contained in $[a, b]$. Thus $I = \bigcup_N [0, g_i]$ is short and k -restrainable in $[0, b]$. By 2.21–22, $I \models B\Sigma_{k+1}$. $\square$

3.21 Corollary. Let $k \geq 1$.

- (1) $I\Sigma_k$ does not prove $(\forall x, z)(\exists y)([x, y] \xrightarrow{*} (k+2)_z^{k+1})$.
- (2) The functions $F_{k,n}(x) = \min_y([x, y] \xrightarrow{*} (k+2)_n^{k+1})$ ($n = 1, 2, \dots$) form an envelope for $I\Sigma_k$ -provably recursive functions.

Proof. (1) follows immediately from 3.10 (i).

(2) follows easily from 3.10 (iv). □

3.22 Remark ($I\Sigma_1$). (a) Recall II.2.16: if $[a, b] \xrightarrow{*} (e+1)_d^e$ and $d = c^3$ then $[a, b] \xrightarrow{**} (e+1)_c^e$.

(b) Observe that II.2.12, together with the fact that for $c \geq 7$ we have $c \geq 1 + 2\sqrt{c}$, implies the following: $c \geq 7$ and $[a, b] \xrightarrow{*} (e+2)_c^{e+1}$ implies $[a, b] \xrightarrow{*} (e+1)_c^e$.

3.23 Theorem. The function

$$Y(a, b) = \max_c([a, b] \xrightarrow{*} (c+1)_c^c)$$

is an indicator for models of PA in $I\Sigma_1$.

Proof. If $a < b \in M \models I\Sigma_1$, $a < I < b$, $I \subseteq_e M$ and $I \models PA$ then obviously, for each standard $n > 0$, $I \models (\exists y)([a, y] \xrightarrow{*} (n+1)_n^n)$; thus $Y(a, b)$ is non-standard.

Conversely, let $Y(a, b) = d$ be non-standard; let u be non-standard such that $e = 2^{6u} \leq d$. By 3.22, $[a, b] \xrightarrow{*} (u+2)_e^{u+1}$ and thus $[a, b] \xrightarrow{*} (u+2)_c^{u+1}$ for $c = 2^{2u}$. Apply 3.17: in $[a, b]$, there are g, h of length u such that g is a Paris sequence for h in b w.r.t. df_1 , and $I = \bigcup_N [0, g_n]$ is short in $[a, b]$. By 3.14 (2), it is k -restrainable in $[0, b]$ for each $k \in N$ and consequently $I \models PA$. □

3.24 Corollary. (1) PA does not prove

$$(\forall x, u)(\exists y)([x, y] \xrightarrow{*} (u+2)_{u+1}^{u+1}).$$

The functions

$$F_n(x) = \min_y([x, y] \xrightarrow{*} (n+2)_{n+1}^{n+1})$$

form an envelope for PA -provably recursive functions.

(d) Paris Sequences of the Second Kind

Here we exhibit another indicator for models of $B\Sigma_{k+1}$ and of PA , using the notion of α -large sets and the Schwichtenberg-Wainer hierarchy.

3.25 Definition ($I\Sigma_1$). Let $g < b$ and let par be an increasing sequence of elements of $[0, b]$. $c \in df_2(b, g, par)$ if there is a formula $\varphi(x, y, z) \in L'$ and parameters e evaluating y such that

- (1) $\langle \varphi, e \rangle < g - 1$, and
- (2) c is the minimal element of $[0, b]$ satisfying $\varphi(x, e, par)$ in $[0, b]$.

3.26 Remark. Compare this with 3.16 (definition of df_1): there we had two independent bounds for the formula and parameters distinct from par . Here we have one point bound for the pair formula-parameters. The main property of df_2 is that, for each b, g, par , the cardinality of $df_2(b, g, par)$ is less than g . (And clearly, df_2 is Δ_1 -defined.)

3.27 Lemma ($I\Sigma_1$). Assume $A \subseteq [a, b]$ to be ω_u^v -large, where $u, v \geq 1$, $a \geq 2$ and let par be an increasing sequence of elements of $[0, b]$. Then there is an $A_1 \subseteq A$ and a $h \in A$, $h > \max A_1$ such that

- (1) A_1 is ω_{u-1}^{v-1} -large and
- (2) if $\{g_i \mid i < \nu\}$ is the increasing enumeration of A_1

then

$$[(g_{i+1}, h)] \cap df_2(b, g_i, par) = \emptyset$$

for each $i < \nu - 1$. (Remark: $[(x, y)]$ denotes the open interval with endpoints x, y .)

Proof. Let $g_0 = \min A$, $h_0 = \max A$. We construct a sequence of intervals $[g_i, h_i]$ ($i = 0, \dots, \nu$) such that $g_i < g_{i+1} < h_{i+1} \leq h_i$ as follows: let $A \cap [g_i, h_i]$ be $\omega_{u-1}^{v-1}(g_0-1)\dots(g_{i-1}-1)$ -large and assume the exponent $\{\omega_{u-1}^v(g_0-1)\dots(g_{i-1}-1) = \alpha$ to be ≥ 2 . Then use the increasing enumeration of $df_2(b, g_i, par)$ to define a decomposition of $[g_i, h_i]$ onto $\leq g_i$ intervals $[x, y]$ such that $[(x, y)] \cap df_2(b, g, par) = \emptyset$. Apply II.3.21 (4) to obtain a $[g_{i+1}, h_{i+1}] \subseteq [(g_i, h_i)]$ such that $A \cap [g_{i+1}, h_{i+1}]$ is $\omega_{u-1}^{\alpha}(g_i-1)$ -large and $[g_{i+1}, h_{i+1}] \cap df_2(b, g_i, par) = \emptyset$.

Let u be the least number such that $\{\omega_{u-1}^v(g_0-1)\dots(g_{\nu-1}-1) \leq 1$. Take a $g_\nu (= g_{\nu-1} + 1, \text{ say})$; then $\{g_i - 1 \mid i \leq \nu\}$ is ω_{u-1}^v -large, i.e. $\{g_i - 1 \mid 1 \leq i \leq \nu\}$ is $\{\omega_{u-1}^v(g_0-1)\dots(g_{\nu-1}-1)\}$ -large, thus $\{\omega_{u-1}^v\}(1)$ -large and therefore ω_{u-1}^{v-1} -large. But then also $\{g_i \mid 0 \leq i \leq \nu - 1\}$ is ω_{u-1}^{v-1} -large. Put $h = h_{\nu-1}$; then the assertion of the lemma holds. This completes the proof. $\square$

3.28 Lemma. There is an m such that $I\Sigma_1$ proves the following: if $a \geq \overline{m}$ and assumptions of 3.27 hold then $g_{i+1} > 2^{g_i}$ for all $i \leq \nu - 2$.

Proof. If $a > \overline{m}$ for an appropriate m then for each g_i there is an $x < g_i < 2^x$ such that $2^{2^x} \in df_2(b, g_i, \emptyset) \subseteq df_2(b, g_i, par)$. (See the sublemma

below.) Now assume $g_{i+1} \leq 2^{g_i}$; then $g_{i+1} \leq 2^{2^x}$, thus $h_{i+1} \leq 2^{2^x}$ (since $[(g_{i+1}, h_{i+1})] \cap df_2(b, g_i, par) = \emptyset$).

Case 1. $g_{i+1} < 2^x$, then $h_{i+1} \leq 2^x < 2^{g_i} < 2^{g_{i+1}}$.

Case 2. $2^x \leq g_{i+1}$, then $h_{i+1} \leq 2^{2^x} \leq 2^{g_{i+1}}$.

In both cases, $h_{i+1} < f_2(g_{i+1}) - 1$ (f_2 is the second function of the Schwichtenberg-Wainer hierarchy, cf. II.3.31), thus $[g_{i+1}, h_{i+1}]$ is not ω^2 -large. Thus $\alpha = \{\omega_{u-1}^v\}(g_0 - 1) \dots (g_i - 1) \leq 1$ (since $[g_{i+1}, h_{i+1}]$ is ω^α -large and $\alpha \geq 2$ implies $\alpha \xrightarrow{*} 2$). We get $i + 1 = \nu - 1$ (g_{i+1} is the last element of g). It remains to prove a sublemma to complete our proof.

3.29 Sublemma. There is an m such that $I\Sigma_1$ proves the following: if $b > g > \overline{m}$ then there is an $a < g < 2^a$ such that $2^{2^a} \in df_2(b, g, \emptyset)$.

Proof. Let $\varphi(x, y)$ define $y = 2^{2^x}$; for the standard pairing function we have $\langle \varphi, a \rangle = a^2/2 + ka + h$ for some k, h ; find m such that $a > m$ implies $ka + h < a^2/2$, then $\langle \varphi, a \rangle < a^2$. Further, let $m \geq 37$ which guarantees that for $g \geq m$ there is an a such that $a^2 < g < 2^a$. (This works for g even; for g odd the proof is analogous.) □

3.30 Lemma ($I\Sigma_1$). Let $a \geq 2$, $A \subseteq [a, b]$ ω_u^v -large; then there is an increasing sequence h of elements of $[a, b]$ with $lh(h) = u$ and an increasing sequence g of elements of A with $lh(g) = v - u$ such that g is a Paris sequence for h in b w.r.t. df_2 .

Proof. Apply iteratively Lemma 3.27: put $A_0 = A$, $par_u = \emptyset$; you get $A_1 \subseteq A$ ω_{u-1}^{v-1} -large and a $h_{u-1} \in A$, $h_{u-1} > A_1$ such that for two consecutive elements x, x' of A_1 we have

$$[(x', h_{u-1})] \cap df_2(b, x, par_u) = \emptyset.$$

Put $par_{u-1} = \{h_{u-1}\}$. If we have $A_i \subseteq A_{i-1}$, A_i ω_{u-i}^{v-i} -large, $h_{u-1}, \dots, h_{u-i} > A_i$, and $par_{u-1} = (h_{u-i} \dots h_{u-1})$ such that $x, x' \in A_i$, $x < x'$ implies $[(x', h_{u-i})] \cap df_2(b, x, par_{u-i+1}) = \emptyset$, then we can use 3.27 to get the same with i replaced by $i + 1$. Finally we get A_u ω_0^{v-u} -large (i.e. $(v - u)$ -large, thus having at least $v - u$ elements) and $par_u = (h_0, \dots, h_{u-1}) = h$ such that if $g = \langle g_i \mid i < v - u \rangle$ enumerates A_u (or: the first $v - u$ elements of A_u) then g is a Paris sequence for h in b w.r.t. df_2 . □

3.31 Remark. Note that by 3.28 if we assume $a \geq \overline{m}$ (where m is as in 3.28) then we may conclude that the sequence grows exponentially.

3.32 Lemma. Let $M \models I\Sigma_1$ be non-standard and let a, g, h, b be as in 3.30–3.31. Assume $v - u$ non-standard and let $J \subseteq v - u$ be a cut; put $I = \sup\{g_i \mid i \in J\}$ (cf. 3.13). Then df_2 is suitable for g, h, b, I .

Proof. Cf. the proof of 3.19. We verify that for $B_j = \bigcup_{i \in J} df_2(b, g_i, par_j)$, $par_j = (h_j, \dots, h_{u-1})$, B_j is closed under Skolem functions. Let $a_1, \dots, a_p \in B_j$ ($p \in N$), let φ be standard and let $c \in [0, b]$ be defined in $[0, b]$ by φ from $a_1, \dots, a_p$. Assume $a_1 \dots a_p \in df_2(b, g_i, par_j)$; let a_i be defined by $\langle \varphi_i, e_i \rangle < g_q$ ($q \in J$, φ_i not necessarily standard).

Thanks to exponential growth of g , there is a $q' = q + k$ for some standard k such that the formula φ' saying

$$(\exists x_1, \dots, x_p) (\bigwedge_i \varphi_i(x_i, y_i, par_j) \& \varphi(z, x))$$

together with the juxtaposition e' of $e_1, \dots, e_p$ form a pair $\langle \varphi', e' \rangle < g_q$; clearly, φ' defines c from e' , par_j in $[0, b]$. This shows that B_j is closed under definable Skolem functions. $\square$

3.33 Theorem. For each $k \geq 1$, the function

$$\begin{aligned} Z_k(a, b) &= \max_c([(a, b)] \text{ is } \omega_k^c\text{-large}), \text{ or} \\ &= 0 \text{ if such } c \text{ does not exist,} \end{aligned}$$

is an indicator both for models of $B\Sigma_{k+1}$ and for models of $I\Sigma_k$ in $I\Sigma_1$.

Remark. This is a well-defined Δ_1 function in $I\Sigma_1$ since $a + 1 > 0$ and therefore $[a + 1, b - 1]$ is not ω_k^{b-a} -large (prove this using II.3.21 (5)). There is a maximal c such that $[(a, b)]$ is ω_k^c -large, since the property in question is a Δ_1 property of c .

Proof. Let $M \models I\Sigma_1$, M non-standard, and let $a < b \in M$.

(a) Similarly as in 3.20 we show that if there is a cut $a < I < b$ such that $I \models I\Sigma_k$ then $Z_k(a, b)$ is non-standard.

(b) Conversely, assume $c = Z_k(a, b) > N$ in M . Then $b > N$ (by the remark above) and we may also assume $a > N$ (otherwise take N for I). Apply 3.30: $[a + 1, b - 1]$ is ω_k^c -large, therefore there are $h = (h_0, \dots, h_{k-1})$ in $[0, b]$ and a Paris sequence g for h w.r.t. df_2 , g of length $c - k > N$. By 3.31, g grows exponentially. Thus by 3.14, $I = \bigcup_n [0, g_n]$ is k -restrainable in $[0, b - 1]$ and therefore $I \models B\Sigma_{k+1}$. $\square$

3.34 Corollary. For $k \geq 1$, the functions $f_{\omega_{k-1}^n}$ ($n = 1, 2, 3, \dots$) form an envelope for $I\Sigma_k$ -provably total Σ_1 functions.

Proof.

$$\begin{aligned} g_n(x) &= \min_y (Z_k(x, y) \geq \bar{n}) \\ &= \min_y (\max_c([(x, y)] \text{ is } \omega_{k-1}^c\text{-large}) \geq \bar{n}) \\ &= \min_y ([x, y] \text{ is } \omega_{k-1}^n\text{-large}) = \min_y ([x + 1, y] \text{ is } \omega_{k-1}^n\text{-large}) \\ &= f_{\omega_{k-1}^n}(x + 1). \end{aligned} \quad \square$$

3.35 Corollary. Let $k \geq 1$; then $I\Sigma_k$ does not prove f_{ω_k} to be total.

Proof. By the indicator theorem $I\Sigma_k$ does not prove $(\forall x, z)(\exists y)([x, y] \text{ is } \omega_k^z\text{-large})$. By II.3.34, this means that $I\Sigma_k$ does not prove that each z , $f_{\omega_{k-1}^z}$ is total. But this means that $I\Sigma_k$ does not prove f_{ω_k} to be total. $\square$

3.36 Theorem. The function

$$\begin{aligned} Z(a, b) &= \max_c([a, b] \text{ is } \omega_c^{2c}\text{-large}) \\ &= 0 \text{ if such } c \text{ does not exist} \end{aligned}$$

is an indicator for models of PA in $I\Sigma_1$.

Proof. Clearly, if $a < I < b$ and $I \models PA$ then for each u ,

$$I \models (\exists y)([a, y] \text{ is } \omega_u^{2u}\text{-large}).$$

Therefore $Z(a, b) > N$. Conversely, if $c = Z(a, b) > N$ then assume $a > N$; by 3.30, there are g, h (h of length c) such that g is a Paris sequence of the second kind for h of non-standard length c . By 3.31, g has exponential growth. Thus by 3.14, there is an $I \subseteq_e M$, such that for each k I is k -restrainable in $[0, b-1]$. Thus $I \models PA$. $\square$

(e) Further Consequences

3.37 Theorem. For each k , $I\Sigma_1$ proves that $(W)_k$ implies $(PH)_k$. (This is Theorem II.3.36; $(W)_k$ is $(\forall x, z)(\exists y)([x, y] \text{ is } \omega_k^z\text{-large})$ and $(PH)_k$ is the k -th instance of the Paris-Harrington principle:

$$(\forall x, z, q)(\exists y)([x, y] \xrightarrow{*} (q)_z^{k+1}).$$

Proof. Let $M \models I\Sigma_1$ be non-standard satisfying $(W)_k$, let $k \geq 1$, $a \in M - N$, $c \in M - N$; we want to find a b such that $M \models [a, b] \xrightarrow{*} (k+2)_c^{k+1}$. This will clearly verify $M \models (PH)_k$. We may assume $M \models c < a$ (otherwise change a to c). Find a b such that $M \models [a, b]$ is ω_k^c -large where $e = c^\nu + k$ for some non-standard ν and put $d = 2^{2^b}$; in M , $[a, d]$ is also ω_k^e -large.

Let $f \in M$, $f : [a, b]^k \rightarrow c$; then clearly f is coded in $[0, d]$. Apply 3.30 and 3.31 with b replaced by d and with A replaced by $[a, b]$; find g, h in $[0, d]$ such that $lh(h) = k$, $lh(g) = c^\nu$ and g is an exponentially growing Paris sequence for h w.r.t. df_2 , consisting of elements of $[a, b]$.

Claim. There is a cut $J < c^\nu$ such that $I = \sup_J \{g_i \mid i \in J\}$ is regular up to c .

Proof of the claim. Enumerate all coded partitions of the set $[0, b]$ into c parts as $\bigcup_{j < c} q_{ij}$, $i \in N$ (this system exists due to the countability of M). Define a sequence $\{A_i \mid i \in N\}$ such that $A_0 = A$, for each i , $\text{card}(A_i) \geq c^{v-i}$ and for some j , $A_{i+1} \subseteq A_i \cap q_{ij}$. Then put $I = \sup_N \min(A_i)$. I is a cut in $[0, d]$ and if $\bigcup_{j < c} d_j$ is a coded partition of I then it induces a partition $\bigcup_j q_{ij}$ for some i ($d_j = q_{ij} \cap I$). Take A_{i+1} and let $A_{i+1} \subseteq q_{ij_0}$. Then clearly $A_{i+1} \cap I$ is unbounded in I , thus d_{j_0} is unbounded in I . This proves the claim.

We continue the proof of the theorem. Apply 3.14. Thus we have a chain

$$a \in I \subseteq_e K_0 \prec_I \cdots \prec_I K_k \prec [0, d];$$

by inspecting the proof of 3.30 we see that we may assume $f \in K_0$. (Put $\text{par}_u = (f)$.) Observe that $b \notin I$ and apply 2.24: we have $K_0 \models [a, b] \rightarrow^* (k + 2)_c^{k+1}$ and $f \in K_0$, thus in K_0 and consequently in M , f has a relatively large homogeneous set. This completes the proof. $\square$

3.38 Corollary. For each k , the following are equivalent over $I\Sigma_1$:

- (i) $\text{Con}(I\Sigma_k + \text{Tr}(\Pi_1))$, (ii) $(PH)_k$, (iii) $(W)_k$.
- (See II.3.36.)

3.39 Theorem. Each non-standard model of $I\Sigma_1$ has a non-standard cut that is a model of PA .

Proof. For each standard u , $N \models (\exists y)(Y(u, y) \geq u)$, thus if M is non-standard and $M \models I\Sigma_1$ then $M \models (\exists y)(Y(u, y) \geq u)$. (Y is the indicator for PA in $I\Sigma_1$.) By overspill, there is a non-standard $c \in M$ such that $M \models (\exists y)(Y(c, y) \geq c)$. Let d be such a y ; then in M , $Y(c, d)$ is non-standard and therefore is a cut $c < I < d$, $I \models PA$. $\square$

4. Formalizing Model Theory

The main aim of this short section is to exhibit a proof of Paris's conservation result (for all n , $B\Sigma_{n+1}$ is a Π_{n+2} -conservative extension of $I\Sigma_n$, see Sect. 1) in $I\Sigma_1$. This implies that for each n $I\Sigma_{n+1}$ proves $\text{Con}^*(B\Sigma_{n+1}^*)$ (since it proves $\text{Con}^*(I\Sigma_n^*)$, see I.4.34). This is a valuable piece of information on the relation of fragments of PA ; for example, it implies that $I\Sigma_{n+1}$ is not interpretable in $B\Sigma_{n+1}$.

Thus the section is a free continuation of Chap. I, Sect. 4 where we formalized some parts of logic in $I\Sigma_1$. Our present formalization consists in formalizing Paris's proof of his conservation result; roughly, we show in

$I\Sigma_1$ that each model of $I\Sigma_n$ has a satisfactorily elementary extension to a model of $B\Sigma_{n+1}$ and that this implies the desired conservation. Thus we shall see that important parts of the model theory of fragments can be developed inside $B\Sigma_2$. Central devices are the Low arithmetized completeness theorem (cf. I.4) and a consistency criterion (see below). The section has the following structure: (a) some results on satisfaction and consistency, (b) Paris's conservation result in $B\Sigma_2$, (c) a proof of conservativity of ACA_0 over PA .

(a) Some Results on Satisfaction and Consistency

Recall Chap. I, Sect. 4 (c); there we formalized model theoretical notions in $I\Sigma_1$ and after having proved the Low arithmetized completeness theorem (I.4.27) we investigated (inside $I\Sigma_1$) the language of arithmetic and defined its standard model (the universe). For each k , $I\Sigma_1$ proves that the standard model has a satisfaction for $(\Sigma_k \cup \Pi_k)$ -formulas; but it follows easily by self-reference that $I\Sigma_1$ proves the non-existence of a full satisfaction for the standard model (for a precise statement see below).

Now that we have studied non-standard models of fragments, we are ready to continue the development of positive results in fragments by investigating, in some appropriate fragment, non-standard models of arithmetic. The Low arithmetized completeness theorem yields models with full satisfaction; but other constructions need not. We present here a theorem enabling us to derive, in $I\Sigma_1$ (or in $I\Sigma_1(\Gamma)$ for some Γ), the consistency from the existence of a model with partial satisfaction. This will be used in the next subsection during our proof of Paris-Friedman's conservation theorem in $I\Sigma_1$.

4.1 Theorem. For each k , $I\Sigma_1$ proves the following: there is no $Sat \in \Sigma_k^*$ such that Sat is a full satisfaction for the standard model. (Cf. I.4.23 for the notion of a full satisfaction.)

Proof. Let $\sigma(z, e, p)$ be a Σ_k formula defining binary Σ_k^* -relations in $I\Sigma_1$; let, by the diagonal lemma III.2.1, $\varphi(p)$ be such that $I\Sigma_1 \vdash \varphi(p) \equiv \neg\sigma(\overline{\varphi(p)}, [p], p)$. Now assume in $I\Sigma_1$ that we have a p such that $Sat = \{\langle z, e \rangle \mid \sigma(z, e, p)\}$ is a full satisfaction class for the universe. Then on the one hand, $\varphi(p)$ is equivalent to $\neg\sigma(\overline{\varphi(p)}, [p], p)$, but on the other hand, by "it's snowing"-it's snowing applied to φ , $\varphi(p) \equiv \langle \varphi(p), [p] \rangle \in Sat \equiv \sigma(\varphi(p), [p], p)$, which is a contradiction. $\square$

4.2 Convention. In the rest of the subsection we assume that T is a theory extending $I\Sigma_1$ and Γ is a class of formulas such that $T \vdash \Delta_1(\Gamma) \subseteq \Gamma$ and $T \vdash I\Sigma_1(\Gamma)$. Our favourite choice will be $\Gamma = LL_1$ (i.e. low $\Sigma_0^*(\Sigma_1)$) and

$T = I\Sigma_1$, cf. Chap. I, Sect. (c). Instead of saying that we work in T , just say that we work in $I\Sigma_1(\Gamma)$.

4.3 Theorem ($I\Sigma_1(\Gamma)$). If L is the language of arithmetic and (M, S, Sat) is a Γ -defined model for L where Sat is a satisfaction for (M, S) and $(\Sigma_k^\bullet \cup \Pi_k^\bullet)$ -formulas then the Π_{k+1} theory of (M, S) is consistent.

Proof. Note that Sat uniquely extends to a satisfaction Sat' for Π_{k+2} formulas; but Sat' need not be Γ . The Π_{k+2} theory of (M, S) is the set of all Π_{k+2} sentences true in (M, S) . The theorem follows by I.4.26 (relativized to $I\Sigma_1(\Gamma)$) from the following lemma: $\square$

4.4 Lemma ($I\Sigma_1(\Gamma)$). If (M, S, Sat) is as in 4.3 then, for each finite set T_0 of Π_{k+2} -formulas true in (M, S, Sat) , (M, S) can be expanded to a Γ -defined model (M, S') of skolemizations of formulas from T_0 (with respect to the satisfaction guaranteed by I.4.24).

Proof. Observe that theorem I.4.33, stating that for $n \geq 1$, $I\Sigma_n$ proves the consistency of the set of all true $\Pi_{n+1}^\bullet$ sentences, is a particular case of our present theorem 4.3: $T = I\Sigma_n$, $\Gamma = \Delta_n$, $T \vdash I\Sigma_1(\Gamma)$, (M, S) is the standard model* with the Δ_n satisfaction for $(\Sigma_{n+1}^\bullet \cup \Pi_{n+1}^\bullet)$ formulas*. Thus take $k = n - 1$: we get $T \vdash Con^*(Tr(\Pi_{k+2}^\bullet))$, and $k + 2 = n + 1$. $\square$

The proof of I.4.33 is just a particular case of the proof of our present lemma; the reader may generalize easily. (Hint: all least number operators are understood as taken outside the model, not in the sense of the model. Values of terms are well defined since all function symbols in question are interpreted as Γ -functions.)

4.5 Corollary. $I\Sigma_1$ proves that if $M = (M, S, Sat)$ is an LL_1 model and Sat is a satisfaction for $(\Sigma_k^\bullet \cup \Pi_k^\bullet)$ formulas ($k \geq 1$) then the $\Pi_{k+2}^\bullet$ theory of M is consistent.

Proof. By 4.3 using our favourite choice (cf. 4.2). $\square$

(b) A Conservation Result in $I\Sigma_1$

We formalize the considerations of Sect. 1 (f); in particular, we formalize the proof of 1.61 in $I\Sigma_1$.

4.6 Theorem ($I\Sigma_1$). Let $k \geq 0$ and let $(I\Sigma_k^\bullet)^*$ be the strong Skolem extension of $I\Sigma_k^\bullet$ as described in I.4.10. Let M^* be an LL_1 model of $(I\Sigma_k^\bullet)^*$ with full satisfaction and let M be the reduct of M^* to the language of arithmetic.

Then there is an LL_1 model I of $B\Sigma_{k+1}^\bullet$ with satisfaction for $\Sigma_{k+1}^\bullet \cup \Pi_{k+1}^\bullet$ -formulas such that I is a cofinal, $(k+1)$ -elementary extension of M .

4.7 Corollary. $I\Sigma_1$ proves that, for each $k \geq 0$, $B\Sigma_{k+1}^\bullet$ extends $I\Sigma_k^\bullet \Pi_{k+2}^\bullet$ -conservatively. (See the remark following 1.61 and use 4.5 with $(k+1)$ instead of k).

4.8 Corollary. For each $k \geq 1$, $I\Sigma_k$ proves $\text{Con}^\bullet(B\Sigma_k^\bullet)$. (Since, by $I\Sigma_k \vdash \text{Con}(I\Sigma_{k-1}^\bullet)$ and, furthermore, $I\Sigma_k \vdash \text{Con}^\bullet(I\Sigma_{k-1}^\bullet) \rightarrow \text{Con}^\bullet(B\Sigma_k^\bullet)$.)

In the rest of the subsection, we elaborate the proof of 4.6.

4.9 Construction. We proceed in $I\Sigma_1$. Our aim is to formalize the proof of 1.61. Let M^* be as in 4.6 above. We shall carefully define an ultrapower K^* of M^* using an ultrafilter U of definable subsets of M^* . The care concerns not only the verification that our construction works in $I\Sigma_1$, but also that our special choice of U will ensure that the cut $I = \sup_K(M)$ (where K, M are reducts of K^*, M^* to the language of arithmetic) will be Δ_1 in M^* , i.e. will itself be LL_1 . We shall define low Δ_2 satisfaction for K^* using Łoś's theorem and define an LL_1 satisfaction on I for $\Sigma_k \cup \Pi_k$ -formulas such that $I \prec_k K$. An additional proof will show that even satisfaction for $\Sigma_{k+1} \cup \Pi_{k+1}$ on I is $\Delta_1(M)$, i.e. LL_1 . The reader is advised to compare our construction with 1.65.

(1) K_0 will be the set of all parametrical definitions of mappings of M^* into itself, i.e. of formulas $\varphi(y, x, a)$ of the language of $(I\Sigma_k^\bullet)^*$ such that $M^* \models (\forall x)(\exists! y)\varphi(x, y, a)$ (a is a parameter from M). Clearly, $K_0 \in \Delta_1(M^*)$ (here M^* denotes the structure together with its full satisfaction) and hence $K_0 \in LL_1$. Similarly, B will denote the set of all parametrical definitions of subsets of M , i.e. of arbitrary formulas $\varphi(x, a)$. Again, $B \in \Delta_1(M^*)$.

Since $I\Sigma_1$ proves $I\Sigma_1(LL_1)$ and both K_0 and B are unbounded, we may work with increasing enumerations of both sets; the enumerations are $\Delta_1(M^*)$. Thus let

$$\begin{aligned} K_0 &= \{f_1, f_2, \dots\} \\ B &= \{\varphi_1, \varphi_2, \dots\} \end{aligned}$$

(2) We shall construct a "generic sequence" of elements of B defining an ultrafilter U . In the even steps, we shall deal with φ_i 's and in the odd steps we shall decide the membership to the cut I . Using $I\Sigma_1(LL_1)$ we construct by induction a sequence of formulas $\alpha_j(x)$ (containing parameters from M^*); for each j , $\Omega_j(x)$ abbreviates $\bigwedge_{z < j} \alpha_z(x)$. Put $\alpha_0(x) = \text{TRUE}$.

Step $j = 2i + 1$. Let $\Phi \equiv (\exists y)(0x)(\Omega_j(x) \& f_i(x) \leq y)$.

Case 1. $M^* \models \Phi$. Then let y_0 be the least element of M^* (in the ordering of the universe) such that $M^* \models (Cx)(\Omega_j(x) \& f_i(x) \leq y_0)$ and put $\alpha_j(x) = f_i(x) \leq y_0$.

Case 2. Otherwise let $\alpha_j(x) = \text{TRUE}$.

Step $j = 2i + 2$. If $M^* \models (Cx)(\Omega_j(x) \& \varphi_i(x))$ then $\alpha_j(x)$ is $\varphi_i(x)$, otherwise $\alpha_j(x)$ is $\neg\varphi_i(x)$.

Let us stress again that the sequence of α 's is $\Delta_1(M^*)$ and hence LL_1 ; for each j we have $M^* \models (Cx)\alpha_j(x)$. Let U be the set of all φ_i such that in step $2i + 2$ we have taken $\alpha_j(x)$ to be $\varphi_i(x)$; again $U \in \Delta_1(M^*)$.

(3) For $f, g \in K_0$ put $f =_U g$ if the formula $f \equiv g$ belongs to U (in other words, the set $\{x \in M \mid M \models f(x) = g(x)\}$ is in U). We define $f + g$ and $f \cdot g$ "coordinate-wise", i.e. $f + g$ is defined as the function associating with each x the value $f(x) + g(x)$ etc. Similarly for $0, S, \leq$. Observe that $=_U$ is a congruence for all these. Thus we have defined K_0 as a structure for the language of arithmetic with non-absolute equality. We may convert K_0 to the structure K with absolute equality by selecting from each equivalence class of $=_U$ its least element. Clearly, $K \in \Delta_1(M^*)$.

(4) We define the full satisfaction for K as follows:

$$K \models \varphi(f_1, \dots, f_n) \text{ iff } \{x \in M \mid M \models \varphi(f_1(x), \dots, f_n(x))\} \in U.$$

This definition is $\Delta_1(M^*)$: recall that $f_1, \dots, f_n$ are formulas of $(I\Sigma_1)^*$ and so is $\varphi(f_1(x), \dots)$. In the above enumeration, the last formula is φ_i , say. Thus the left hand side above holds iff in step $2i + 2$ of the construction of U , the first case occurs (and φ_i is put into U). We are obliged to prove that this is indeed a satisfaction:

Lemma. The definition above defines full satisfaction on K , i.e. Tarski's conditions hold true. (Easy.)

(5) We embed M into K by identifying each $a \in M$ with the constant function on M with the value a (i.e. with the formula $y = a$). Then the definition of satisfaction for K immediately gives $M \prec K$. The proof of the fact that M is not cofinal in K is as in 1.66.

(6) Define

$$I = \sup_K(M) = \{f \in K \mid (\exists y \in M)(\{x \mid f(x) < y\} \in U)\}.$$

By this definition, I is $\Sigma_1(M^*)$; but it is even $\Delta_1(M^*)$ since $f \in I$ if for the i such that $f = f_i$ (in the enumeration of K_0 above) we have Case 1 in step $2i + 1$. Thus I is low Δ_2 .

(7) Now define satisfaction for $\Sigma_k^\bullet \cup \Pi_k^\bullet$ -formulas on I using the satisfaction on K , i.e. put

$$I \models \varphi(a) \text{ iff } K \models \varphi(a)$$

for $a \in I$ and $\varphi \in \Sigma_k^\bullet \cup \Pi_k^\bullet$; use the proof of 1.66 (3) to show that I is indeed a satisfaction, i.e. Tarski's conditions hold true.

(8) Extend this to a satisfaction $\models'$ on I for $\Sigma_{k+1}^\bullet \cup \Pi_{k+1}^\bullet$. We show that $\models'$ is also $\Delta_1(M^*)$. It suffices to show that satisfaction for $\Sigma_{k+1}^\bullet$ is $\Delta_1(M^*)$ so let $\models'$ mean this satisfaction. Clearly, $\models'$ is $\Sigma_1(M^*)$; we show that it is also

$\Pi_1(M^*)$. We know that K is a model of $I\Sigma_k$ (since $M \prec K$ and $M \models I\Sigma_k$) and I is a cut in K . By underspill in K , for each Σ_{k+1}^* -formula $(\exists x)\varphi(x, y)$,

$$I \models \neg(\exists x)\varphi(x, a) \text{ iff } (\exists b \in K - I)(K \models \neg(\exists x < b)\varphi(x, a)),$$

which is $\Sigma_1(K, I)$ and therefore $\Sigma_1(M^*)$.

(9) Now we may prove $M \prec_{k+1} I$ exactly as in 1.66 (2), and $I \models B\Sigma_{k+1}$ as in 1.67 (i.e. in 1.22). This completes the proof of 4.27.

(c) Appendix: Another Conservation Result

We shall sketch a proof of the following

4.10 Theorem. $I\Sigma_1$ proves that ACA_0 is a conservative extension of PA .

Proof. We shall imitate the model-theoretic proof presented in III.1.16. In $I\Sigma_1$ we have to deal with LL_1 sets and a reasonably good satisfaction; this needs some care.

In $I\Sigma_1$, assume $Con^*(PA^*)$ and investigate PA^* – the open Δ_1 theory with a language L^* extending PA^* conservatively and such that for each formula $\varphi(x, y)$ of PA^* there is a PA^* -provably equivalent open PA^* -formula^{*}. Let M be a full LL_1 model of PA^* . Extend M to a model K by adding (codes of) parametrically M -definable subsets of M as sets (in such a way that the set of all such sets is LL_1). Define $K \models X = Y$ iff $M \models (\forall x)(\varphi(x) \equiv \psi(x))$ where X is defined by φ and Y by ψ . Make K into a model with absolute equality by the same trick as in subsection (c). If X is given by $\varphi(x)$ then $K \models x \in X$ means $M \models \varphi(x)$. Define a predicate *Set* ranging over all sets of K , a function $L(X)$ selecting the least element of X if X is non-empty and $A(X, Y)$ selecting the least element of the symmetric difference of X, Y if it is non-empty. Finally, for each open formula $\Phi(x, y, Z)$ of PA^* add a function symbol $F_\Phi(y, Z)$ and the axiom

$$Set(F_\Phi(y, Z) \ \& \ [\Phi(x, y, Z) \equiv x \in F_\Phi(y, Z)]).$$

Observe that this can be done in such a way as to make the resulting model (denoted again by K) LL_1 .

Show that the open theory T_0 of K implies ACA_0 and $Th(M)$ (the theory of M). T_0 is consistent by I.4.26 (relativized; thus $ACA_0 + Th(M)$ is also consistent. This completes the proof. $\square$

All pure mathematics follows formally from
twenty premisses.

(Bertrand Russel,
The Principles of Mathematics)

Part C

Bounded Arithmetic

Chapter V

Bounded Arithmetic

In the previous chapters we have encountered some weak fragments of arithmetic. First we have seen that some properties of numbers can be proved and several concepts can be formalized even in very weak fragments. Often we have used $I\Sigma_1$ as a base theory where a weaker theory was possible. In this chapter we shall consider the possibility of formalizing syntax and some basic combinatorial concepts in weaker fragments. Section 3 will be devoted to this aim. Let us stress that the reason for trying to formalize certain concepts in weak fragments is not only our curiosity. In fact weak fragments are often needed in the study of strong theories. Recall the concept of a definable *cut*, Chap. III, Sect. 3. Cuts in fragments of PA which are closed under $+$ and $*$ define interpretations of bounded arithmetic ($I\Sigma_0$ and some stronger systems), but usually they do not interpret even $I\Sigma_1$. Thus we need weak fragments in order to understand the theory of cuts.

The second reason, which is perhaps the main reason for this research into weak fragments, is their close connection to computational complexity. It is obvious that there is some connection. In $I\Sigma_0$ we have induction for classes of bounded formulae. It is well-known that bounded formulae define easily-computable sets; for instance they are computable in linear space. This is however only a superficial observation. Some deeper connections will be shown in Chap. IV. It seems that more such relations are yet to be discovered.

Some knowledge of complexity belongs to a general background of all working mathematicians. Even so, we have decided to include Sect. 2 as a short introduction in complexity theory, since concepts like the Polynomial Hierarchy are quite often disregarded in textbooks.

Section 1 has a similar introductory role. We shall survey the most important weak fragments and prove some basic results on them. Note that some results on such theories (especially Q and I_{open}) have been proved in Chap. I, Sect. 1.

The last section, Sect. 5, deals with classical metamathematical questions in the context of bounded arithmetic: unprovability of consistency, partial conservativity, interpretability. We will also have to develop some proof theory in Bounded Arithmetic.

In the proofs we shall not favor a particular proof technique; we shall mix proof theory with model theory, since our aim is not the study of one of these fields, but the study of weak fragments. An exception is Sect. 4, where we prefer model theory, since the same subject has been treated using proof theory in another book, [Buss 86, Bounded Arith.]; this section however contains several results not included in Buss's book.

This chapter needs very little from the previous chapters and hence can be read almost independently. It requires only basic acquaintance with fragments of Peano arithmetic, Chap. I, Sects. 1 and 2. In Sect. 5 we shall work with definable cuts, thus we need Chap. III, Sect. 3.

1. A Survey of Weak Fragments of Arithmetic

(a) Fragments of Arithmetic

The weakest fragment that we consider in this book is *Robinson's arithmetic* Q . It was defined in Chap. I, Sect. 1. The fragments of arithmetic which use the language L_0 are extensions of Q by adding a schema of induction for a restricted class of formulae and, possibly, an axiom saying that a certain function is total. Though Q proves only very few interesting statements about numbers, (it does not prove for instance the commutativity of addition), we shall see that quite strong theories are *interpretable* in it, (Sect. 5).

Sometimes it is more convenient to start with a stronger theory. Usually one takes P^- which is just the axiomatization of the *nonegative part of the discretely ordered ring*. These axioms are provable in I_{open} and they are listed in Theorem 1.10, Chap. I, where we only have to add

$$\bar{0} \leq x \ \& \ x \leq \bar{1} \rightarrow .x = \bar{0} \vee x = \bar{1}.$$

The next step in the hierarchy is I_{open} . Several basic properties of natural numbers are provable already in this theory, see Chap. I, Sect. 1. In spite of this fact it is a very weak theory. There are models of I_{open} in which the following equations do have solutions

$$\begin{aligned}(x+1)^2 &= 2(y+1)^2; \\ (x+1)^3 + (y+1)^3 &= (z+1)^3;\end{aligned}$$

i.e. the irrationality of $\sqrt{2}$ and Fermat's last theorem for exponent 3 are not provable. Such independence results are possible, because there are recursively defined models of I_{open} . It seems that the properties such as the property of having recursive models determine the border-line between very weak fragments and other fragments of arithmetic. I_{open} is an interesting

object for research. However the methods needed for it are different from those needed for stronger systems, therefore we shall not consider it here. Let us only mention here that one can prove independence results also for some strengthenings of I_{open} (as it is defined here). Then it is more convenient to define I_{open} to be a ring, not only the nonnegative part of it, and add properties such as *normality* (which means that it is integrally closed in its fraction field and which implies that $\sqrt{2}$ is irrational).

The next natural step in the hierarchy is $I\Sigma_0$; let us recall that this is Q with induction schema for all bounded formulae in language L_0 . First we state two theorems which show that there is an essential difference between the structure of models of $I\Sigma_0$ and I_{open} , since none of these theorems holds for I_{open} .¹ The proofs require some facts which will be proved later.

1.1 Theorem. There is no recursive nonstandard model of $I\Sigma_0$; in fact, in each nonstandard countable model M of $I\Sigma_0$ both $+_M$ and $\ast_M$ are not recursive.

Proof. Let A, B a recursively inseparable pair of subsets of N . Let φ and ψ be Σ_0 formulae such that

$$\begin{aligned} n \in A \equiv N \models (\exists y)\varphi(\bar{n}, y), \\ n \in B \equiv N \models (\exists y)\psi(\bar{n}, y). \end{aligned}$$

Since A and B are disjoint, we have

$$N \models (\forall x, y, z) \neg(\varphi(x, y) \& \psi(x, z)).$$

Hence, for every k ,

$$M \models (\forall x, y, z \leq k) \neg(\varphi(x, y) \& \psi(x, z)).$$

Since this is a Σ_0 formula, by overspill we get some $a \in M$ nonstandard such that $(\forall x, y, z \leq a) \neg(\varphi(x, y) \& \psi(x, z))$ holds in M . Let C be defined by

$$n \in C \equiv M \models (\exists y \leq a)\varphi(\bar{n}, y).$$

Then $A \subseteq C$ and $C \cap B = \emptyset$. Let C' be an initial segment of C . We shall code it by an element c' of M in such a way that b belongs to C' iff the b -th prime divides c' . Clearly this coding can be expressed by a Σ_1 formula. Using a similar overspill argument as above one can show that there is a $c \in M$ which codes the whole C , i.e.

$$n \in C \equiv M \models \bar{p}_n \mid c,$$

¹ Let us note that R. Kaye has defined a fragment of arithmetic for which 1.1 holds, but for which 1.2 fails.

where p_n is the n -th prime. Now assuming that $+_M$ is recursive, we can show that C is recursive: given n , search for $r, s \in M$, $r <_M s$, such that

$$M \models c = \underbrace{s + s + \cdots + s}_{p_n \text{ times}} + r.$$

If $r = 0$, then p_n divides c , hence $n \in C$. If $r \neq 0$, then $n \notin C$. Since this is a contradiction with recursive inseparability of A and B , $+_M$ cannot be recursive. In order to show that $*_M$ is not recursive, code C by 2^c . This is possible, if we take c sufficiently small nonstandard. Then the procedure for testing $n \in C$ will search for $u, v \in M$, $u < v$ such that

$$M \models c = \underbrace{v * v * \cdots * v}_{p_n \text{ times}} * u,$$

and decide according to the truth of $u = \bar{1}$. □

1.2 Theorem. Let M be a countable nonstandard model of $I\Sigma_0$. Then there exists a nonstandard initial segment of M which is a model of PA .

Proof. This is Theorem IV.3.39 strengthened from $I\Sigma_1$ to $I\Sigma_0$. Thus one only needs to check that the proof can be carried out in $I\Sigma_0$. We shall use a different argument. By Theorem IV.3.39, it suffices to prove that every countable model $M \models I\Sigma_0$ has a nonstandard initial segment which is a model of $I\Sigma_1$. We shall construct a semiregular short cut I and apply Theorem IV.2.16 to deduce that it is a model of $I\Sigma_1$. We have to use some properties of $I\Sigma_0$ which will be proved later in Sect. 3 (coding of sequences, definition of the graph of exponentiation).

Let M be a countable model of $I\Sigma_0$. Let $g(x, y)$ be defined as follows

$$\begin{aligned} g(0, y) &= y^y; \\ g(x+1, y) &= g^{y+1}(x, y) \quad (y+1 \text{ iterated } \lambda x(g(x, y))). \end{aligned}$$

One can show that $g(x, y) = z$ can be defined by a Σ_0 formula and the inductive clauses can be proved in $I\Sigma_0$. Let $d \in M$ be a nonstandard number such that 2^{2^d} exists in M . By overspill we can find a nonstandard c such that $g(c, c)$ exists in M . Take an enumeration $f_0, f_1, \dots$ of functions coded in M whose domain is $[0, c]$ such that each function occurs in it infinitely often. Define a nested sequence of intervals

$$[a_0, g(c, a_0)) \supseteq [a_1, g(c-1, a_1)) \supseteq \dots$$

with $a_0 = c$, as follows. Suppose the sequence has been defined up to $[a_i, g(c-i, a_i))$. Decompose this interval into

$$\begin{aligned} [a_i, g(c-i-1, a_i)) \cup [g(c-i-1, a_i), g^2(c-i-1, a_i)) \cup \dots \\ \dots \cup [g^{a_i}(c-i-1, a_i), g^{a_i+1}(c-i, a_i)). \end{aligned}$$

By the Pigeon Hole Principle (we can use it, since f_i is coded in M) one of the intervals in the subdivision does not intersect the range of f_i restricted to the domain $[0, a_i)$. This will be the next interval in the sequence. Let I be the cut determined by $a_0, a_1, \dots$. The construction ensures that it is a cut (the successor function occurs infinitely often) it is semiregular (the range of f_i restricted to a_i is bounded by a_{i+1} in I), nonstandard (since $c \in I$) and short (since $I < d$). $\square$

By the definition of Chap. 0 we allow only variables as the bounds in formulae of Σ_0 . Let f be a function symbol. Extend L_0 by f and let us denote by Σ_0^f the class of bounded formulae in the extended language where we allow the bounds at the quantifiers to be *terms in the extended language* $L_0(f)$. Consider the following two theories:

$$\begin{aligned} I\Sigma_0^f(f) + \text{"}f \text{ is monotone"}; \\ I\Sigma_0(f) + \text{"}f \text{ is monotone"}. \end{aligned}$$

In the first one we extend the induction schema of $I\Sigma_0$ to Σ_0^f formulae, while in the second one we allow induction only for bounded formulae in $L_0(f)$ where the bounds at the quantifiers are only variables. We have already considered a special case with f equal to *exp*.

1.3 Proposition. The theories $I\Sigma_0^f(f) + \text{"}f \text{ is monotone"}$ and $I\Sigma_0(f) + \text{"}f \text{ is monotone"}$ are equivalent.

This proposition shows that it is not important which bounds are allowed at bounded quantifiers. The proof is almost identical with the proof of Lemma I.1.30.

It turns out that the *graphs* of exponentiation and several other important functions are definable by Σ_0 formulae. $I\Sigma_0$ does not prove that exponentiation (defined by such a formula) is a total function (see Theorem 1.4 below), but it does prove the inductive properties of it whenever the values exist. Thus instead of working in $I\Sigma_0(\text{exp})$ (with the inductive properties for *exp*) we can take $I\Sigma_0$ with an axiom saying that *exp* defined by a Σ_0 formula is a total function. Then $I\Sigma_0(\text{exp})$, and hence $I\Sigma_0^{\text{exp}}(\text{exp})$, will be conservative extensions. (These extensions are conservative in a stronger sense: the new function symbol is definable in the subtheory.) The same holds for other monotone functions with Σ_0 definable graphs. This shows that extension of $I\Sigma_0$ by an axiom saying that a certain (Σ_0 definable monotone) function is total has the same effect as an extension of the language. Hence it is natural to consider such theories to be versions of the intuitive concept of *Bounded Arithmetic*.

It is useful to realize that $I\Sigma_0^f(f)$ has a $\Pi_1(f)$ axiomatization: replace the induction schema by the following version

$$(\forall x)(\varphi(\bar{0}) \ \& \ (\forall y < x)(\varphi(y) \rightarrow \varphi(S(y))) \rightarrow \varphi(x))$$

for $\Sigma_0(f)$ formulae and omit the axiom (Q3).

Let us recall some definitions from Chap. III.

- (1) $|x| = \lceil \log_2(x+1) \rceil$;
- (2) $\omega_0(x) = 2x$;
 $\omega_{i+1}(x) = 2^{\omega_i(|x|-1)}$, for $x > 0$ and $\omega_{i+1}(0) = 0$;
- (3) $2_0^x = x$,
 $2_{i+1}^x = 2^{2^x}$.

Further we consider the following axioms:

- (1) $\Omega_i \equiv_{df} (\forall x)(\exists y)(\text{"}y = \omega_{i+1}(x)\text{"})$;
- (2) $Exp \equiv_{df} (\forall x)(\exists y)(\text{"}y = 2^x\text{"})$;
- (3) $Superexp \equiv_{df} (\forall x)(\exists y)(\text{"}y = 2_x^x\text{"})$.

Here we use the quotation marks to denote a Σ_0 formula with the same meaning. At this point it is not obvious that such formulae exist. This will be proved in Sect. 3. In $I\Sigma_0$ we can also prove that ω_{i+1} grows faster than ω_i , 2^x grows faster than each ω_i and 2_x^x grows faster than 2^x . Hence if ω_{i+1} is total, then ω_i must be total too, etc. Thus we get the basic hierarchy of theories

$$I\Sigma_0 \subseteq I\Sigma_0 + \Omega_1 \subseteq I\Sigma_0 + \Omega_2 \subseteq \dots \subseteq I\Sigma_0 + Exp \subseteq I\Sigma_0 + Superexp.$$

The most basic theorem about such systems of Bounded Arithmetic is the following one, which is usually referred to as *Parikh's Theorem*.

1.4 Theorem. Let ψ be a Σ_0^f formula and let π be a Π_0^f formula. Suppose that

$$\begin{aligned} I\Sigma_0^f(f) + \pi &\vdash (\forall x)(f(x) \leq f(x+1)), \\ I\Sigma_0^f(f) + \pi &\vdash (\forall x)(\exists y)\psi(x, y). \end{aligned}$$

Then for some term $t(x)$ of $L(f)$

$$I\Sigma_0^f(f) + \pi \vdash (\forall x)(\exists y \leq t(x))\psi(x, y).$$

Proof. Suppose that the conclusion is false for any term t . Let c be a new constant. Thus for every term t , the following theory is consistent

$$I\Sigma_0^f(f) + \pi + (\forall y \leq t(c))\neg\psi(c, y).$$

Since we can bound any finite set of terms by a single one, (for example $I\Sigma_0^f(f) \vdash t_1(x), \dots, t_n(x) \leq t_1(x) + \dots + t_n(x)$), the following theory is also consistent

$$I\Sigma_0^f(f) + \pi + (\forall y \leq t_1(c)) \neg \psi(c, y) + (\forall y \leq t_2(c)) \neg \psi(c, y) + \dots;$$

here we use an enumeration of all terms of $L(f)$. Let M be a model of this theory and let K be defined by

$$K = \{a \in M \mid M \models a \leq t_i(c), \text{ for some } i\}.$$

Since f is monotone, K is a substructure of M . Since $I\Sigma_0^f(f) + \pi$ has a $\Pi_1(f)$ axiomatization K is a model of it. By the construction we have $K \models (\forall y) \neg \psi(c, y)$ which is a contradiction with the assumption of the theorem. $\square$

The theorem can easily be generalized by relaxing the assumptions about the theory. The essential assumptions are that the theory has a $\Pi_1(f)$ axiomatization and proves some basic properties of terms. The theorem naturally transfers to the systems where f is not a function symbol but it is only defined by a bounded formula and an axiom saying that it is a total function is added.

Let us consider two important examples. For $I\Sigma_0$ this means that every Σ_0 definable provably total function is bounded by a *polynomial*. We shall code sequences (say of 0's and 1's) by binary expansions of numbers; the length of a sequence x is thus $|x|$. Hence any provably total function can increase the length of a sequence only *linearly*. This prevents us defining polynomial time computations in $I\Sigma_0$ and complicates also the formalization of other concepts. It can be easily seen that the axiom Ω_1 just captures the polynomial increase of the lengths. Simple computations yield that for every term t with ω_2 there exists a polynomial p such that

$$x \leq t(y) \rightarrow |x| \leq p(|y|)$$

and *vice versa*, for every polynomial p there exists a term with ω_2 such that

$$|x| \leq p(|y|) \rightarrow x \leq t(y).$$

Hence $I\Sigma_0 + \Omega_1$ is a theory in which it is more comfortable to work. In particular $I\Sigma_0 + \Omega_1$ is stronger than $I\Sigma_0$. More generally we have the following corollary.

1.5 Corollary. Let f and g be Σ_0 definable monotone functions. Suppose g grows faster than any function definable by a term in $L(f)$ (in the standard model). Let $\Omega(f)$, resp. $\Omega(g)$, be formulae saying that f , resp. g , is a total

function for some Σ_0 definition of the graph of f , resp. of g . Suppose that $I\Sigma_0 + \Omega(f)$ proves that f is monotone. Then $I\Sigma_0 + \Omega(f)$ does not prove $\Omega(g)$.

Proof. Let $\varphi(x, y)$ be the Σ_0 definition of $f(x) = y$ and let $\psi(x, y)$ be the definition of $g(x) = y$. (Hence e.g. $\Omega(g)$ is $(\forall x)(\exists y)\psi(x, y)$.) Let π be a formula in $L_0(f)$ defined by

$$\pi \equiv_{df} (\forall x, y)(\varphi(x, y) \equiv f(x) = y).$$

Now suppose that $I\Sigma_0 + \Omega(f)$ does prove $\Omega(g)$. Then also

$$I\Sigma_0^f(f) + \pi \vdash \Omega(g),$$

since π implies $\Omega(f)$. By Theorem 1.4

$$I\Sigma_0^f(f) + \pi \vdash (\forall x)(\exists y \leq t(x))\psi(x, y),$$

for some term in $L_0(f)$. But this is a contradiction with the assumption that g grows faster than any function definable by a term in $L_0(f)$. $\square$

1.6 Corollary. The hierarchy

$$I\Sigma_0 \subseteq I\Sigma_0 + \Omega_1 \subseteq I\Sigma_0 + \Omega_2 \subseteq \cdots \subseteq I\Sigma_0 + Exp \subseteq I\Sigma_0 + Superexp$$

is proper. $\square$

The sentences which separate the levels of this hierarchy are Π_2 . We shall see in Sect. 5 that we can separate $I\Sigma_0 + Exp$ from $I\Sigma_0 + \Omega_n$ by a Π_1 sentence and the same is true for $I\Sigma_0 + Superexp$ and $I\Sigma_0 + Exp$. Such separations remain, however, an open problem for the theories below $I\Sigma_0 + Exp$.

*

Now we turn our attention to hierarchies of theories obtained by restricting the quantifier complexity of the formulae allowed in the schema of induction.

1.7 Definition. (1) E_0 is the class of open formulae in L_0 ;

(2) for $n > 0$, E_n is the class of bounded formulae in L_0 in the prenex normal form where the first quantifier is $\exists$, there are at most n alternations of quantifiers and the bounds at the quantifiers are terms in L_0 (which must not contain variables of the current and the following quantifiers).

Example. $(\exists x \leq a)(\exists y \leq a)(xy = a \ \& \ x > \bar{1} \ \& \ y > \bar{1})$ is an E_1 formula.

The theories IE_i are defined in the usual manner: Q plus the induction schema restricted to E_i formulae. Thus IE_0 is just I_{open} . There is a big gap between IE_0 and IE_1 . In particular Theorems 1.1 and 1.2 hold for IE_1 and there are other reasons to consider IE_1 to be relatively strong. On the other hand we are not able to prove that IE_2 is stronger than IE_1 ; worse than that, we even do not know whether IE_1 is weaker than IS_0 .

These questions are connected with difficult open problems in complexity theory and they are probably very difficult too. We can resolve some of these problems, if we assume some conjectures about complexity classes. This will be shown in Sect. 4. We shall use a different hierarchy of theories there. We shall extend language L_0 to a language L_2 by adding more function symbols, we shall use a stronger theory instead of Q and also we modify the definition of the classes of bounded formulae, (see Sect. 4). In this way we obtain a straightforward connection between the Polynomial Hierarchy (these complexity classes will be defined in the following section) and the hierarchy of bounded formulae.

The language L_2 contains a binary function symbol $\#$ whose intended interpretation is $2^{|x|+|y|}$. Other functions of L_2 can be bounded by polynomials. We have the following estimates

$$x \# y \leq \omega_2(4(x+y)^2) + 1;$$

$$\omega_2(x) \leq (x \# x)^4.$$

Hence each function defined by a term in L_2 can be estimated by a function defined by a term in $L_0(\omega_2)$ and *vice versa*. Thus if we use this language for defining a system of bounded arithmetic we should expect that we obtain something equivalent to $IS_0 + \Omega_1$. This system will be denoted by T_2 . There is an equivalent system S_2 based on a modified schema of induction S_2 . The classes in the hierarchy of bounded formulae in L_2 will be denoted by Σ_i^b , $i = 0, 1, \dots$. The corresponding fragments will be denoted by T_2^i and S_2^i . Thus we have for instance: $IE_i \subseteq T_2^i$. We shall concentrate on subsystems of T_2 and S_2 ; we have mentioned the hierarchy IE_i only for sake of completeness.

* *

In the previous part of the book we have encountered several principles which can be used instead of induction. Let us mention here at least the most important one: *the least number principle*. We know that LS_i (the least number principle for Σ_i formulae) is equivalent to IS_i (induction for Σ_i formulae). We can expect that a similar situation occurs in Bounded Arithmetic, if we replace the Arithmetical Hierarchy by the hierarchy of bounded formulae. For most principles this is really so, for instance *induction* and *the least number principle* are equivalent also for the classes Σ_i^b , for $i \geq 1$.

In Bounded Arithmetic we have also some principles whose counterparts for Σ_i have not been considered, since they are trivially equivalent to induc-

tion. These are the schema *PIND*:

$$\varphi(\bar{0}) \ \& \ (\forall x)(\varphi(\lfloor x/2 \rfloor) \rightarrow \varphi(x)) \rightarrow (\forall x)\varphi(x),$$

and schema *LIND*

$$\varphi(\bar{0}) \ \& \ (\forall y < |x|)(\varphi(y) \rightarrow \varphi(S(y))) \rightarrow \varphi(|x|).$$

The intuitive idea behind these schemata is the following. We can, in a sense, verify the conclusion $\varphi(a)$ using the inductive property of φ in *only* $|a|$ steps. Such a verification procedure is thus polynomial in the size of input a , while it is exponential for the ordinary induction. The schema *PIND* is used to define S_2 and its fragments S_2^i .

Another important schema is the *Pigeon Hole Principle*, see Chap. I, 2.20. For higher fragments, the strength of *PHP* is given by Theorem I.2.23. If we consider this principle for Σ_0 formulae, then it is at least as strong as $I\Sigma_0$. Whether it is derivable in $I\Sigma_0$ is an important open problem. We have some evidence that Σ_0 -*PHP* might be stronger than $I\Sigma_0$; we can prove some mathematical results using Σ_0 -*PHP* which we are not able to derive in $I\Sigma_0$ itself. In particular we have

$$Q + \Sigma_0\text{-}PHP \vdash (\forall x)(\exists p)(p > x \ \& \ \text{"}p \text{ is a prime"}),$$

i.e. Σ_0 -*PHP* proves that there are infinitely many primes, while we are not currently able to derive it in $I\Sigma_0$. If Σ_0 -*PHP* is stronger than $I\Sigma_0$ then a philosophical question arises: *what is a more basic principle, Induction or the Pigeon Hole Principle?*

* * *

We have not exhausted all weak fragments that have ever been considered. You can find more in the literature listed in *Bibliographical Remarks and Further Reading*.

2. A Brief Introduction to Complexity Theory

We shall outline some basic concepts of complexity theory. We shall present a little more than is needed for understanding the following sections, but we shall prove only those results which are in some sense connected with the following text. From the point of view of logic, the most interesting subject in complexity is the quantifier hierarchies such as the Polynomial Hierarchy.

(a) Time and Space Complexity Classes

There are two basic computational models in complexity theory: Turing machines and Boolean circuits. Here we consider only the first one, and also we shall use only one type of it.

By a *Turing machine* we mean an *off-line multitape Turing machine*, which is a finite automaton connected with a finite number of infinite (say in one direction only) tapes. The automaton is also called a *finite control*. The tapes consist of cells on which a symbol of a finite alphabet can be written. The automaton operates heads on tapes. A head can read and rewrite the symbols and move to the next cell (left or right). One of the tapes is distinguished as *the input tape*; the automaton cannot write on this tape. The automaton has a distinguished initial state and a subset of final states. A more up-to-date description would be in terms of a very restricted programming language (instead of the finite automaton) operating with linear arrays of symbols (instead of tapes).

A Turing machine is used to define a language (i.e. a set of strings in a finite alphabet) or a mapping from the set of strings in an alphabet to another such set. The first type is called an *acceptor*, the second type is called a *transducer*. An *acceptor* is a Turing machine which has an input tape and several additional tapes called *work tapes*, and it has a distinguished subset of final states called *accepting states*. A word w is accepted if the machine reaches the accepting state when started in the initial state with w on the beginning of the input tape and with the rest of the tapes blank. The language *accepted* by the machine consists of all words accepted by it. A *transducer* is a similar device except that it has one more distinguished write-only tape. It computes a partial function F in an obvious way: to compute $F(w)$, start with w on the input tape and read $F(w)$ on the output tape when it reaches a final state.

In one step, the machine reads the tapes, rewrites them, moves the heads by one cell and changes the state of the finite control. The *time* of the computation on w is the number of steps needed to reach an accepting state. The *space* is the number of cells used on the work tapes.

At first it may seem that Turing machine time is a very bad approximation of the time needed by actual computers, because the actions of a Turing machine are very restricted. It turns out, however, that Turing machines are quite time efficient due to the fact that they have several tapes. In the case of the space complexity measure this is even more true and having more than one work tape is no advantage.

2.1 Definition. Let $f : N \rightarrow N$ be a function. We say that a language L is *accepted in time* (resp. *space*) $f(n)$, if there exists a Turing machine M which accepts L and such that for every n and any w , $|w| = n$, the time (resp.

space) of M on w is at most $f(n)$. We define

$$\begin{aligned} \text{Time}(f(n)) &= \{L \mid L \text{ accepted in time } f(n)\}, \\ \text{Space}(f(n)) &= \{L \mid L \text{ accepted in space } f(n)\}. \end{aligned}$$

Note that we do not fix the alphabet on work tapes. Hence if we increase the size of the alphabets we can obtain a faster and more space efficient machine. This is the idea behind the following theorem.

2.2 Theorem. (a) For $f(n)$ such that $\lim f(n)/n = \infty$ and $c > 0$,

$$\text{Time}(\max(n, f(n))) = \text{Time}(\max(n, cf(n))).$$

(b) For any $f(n)$ and any $c > 0$,

$$\text{Space}(f(n)) = \text{Space}(cf(n)).$$

Interesting complexity classes are obtained by taking the unions over natural classes of bounding functions.

2.3 Definition.

$$\begin{aligned} \text{LinTime} &= \bigcup_{c \in \mathbb{N}} \text{Time}(cn) && (\text{Linear Time}) \\ P &= \bigcup_{c \in \mathbb{N}} \text{Time}(n^c) && (\text{Polynomial Time}) \\ E &= \bigcup_{c \in \mathbb{N}} \text{Time}(c^n) \\ \text{Exp} &= \bigcup_{c \in \mathbb{N}} \text{Time}(2^{n^c}) && (\text{Exponential Time}) \\ \text{LogSpace} &= \text{Space}(\log n) && (\text{Logarithmic Space}) \\ \text{LinSpace} &= \text{Space}(n) && (\text{Linear Space}) \\ \text{PSpace} &= \bigcup_{c \in \mathbb{N}} \text{Space}(n^c) && (\text{Polynomial Space}) \end{aligned}$$

Note that by Theorem 2.2 we have $\text{LogSpace} = \bigcup_{c \in \mathbb{N}} \text{Space}(c \cdot \log n)$ and $\text{LinSpace} = \bigcup_{c \in \mathbb{N}} \text{Space}(c \cdot n)$.

We have the following trivial inclusions

$$\begin{aligned} \text{LinTime} &\subseteq P \subseteq E \subseteq \text{Exp} \\ \text{LogSpace} &\subseteq \text{LinSpace} \subseteq \text{PSpace} \end{aligned}$$

Below we shall explain why these inclusions are in fact strict. We have also trivially $\text{Time}(f(n)) \subseteq \text{Space}(f(n))$, but there is a nontrivial result which gives more:

2.4 Theorem. For $f(n) \geq n$, $\text{Time}(f(n)) \subseteq \text{Space}(f(n)/\log f(n))$.

In the other direction only the trivial inclusion

$$\text{Space}(f(n)) \subseteq \bigcup_{c \in \mathbb{N}} \text{Time}(c^{f(n)})$$

is known. (An $f(n)$ space bounded machine cannot compute in time longer than $c^{f(n)}$, because it would run into a cycle.) Thus for instance, $\text{LogSpace} \subseteq P$, but it is a famous problem whether the inclusion is strict.

The complexity classes above are defined for languages. We shall talk about the complexity of *subsets of nonnegative integers* using the standard binary coding. We can talk also about the complexity of relations, since we can code k -tuples of words by single words using an additional symbol for separating them. Using transducers instead of acceptors we can define corresponding classes of *functions*. (Note that there is no simple reduction of a function class to the corresponding language class.) We shall not need the function classes corresponding to the classes of languages defined above except for the function class corresponding to P ; it will be denoted by $\square_1^P$.

Let us mention an alternative notation: the subsets of natural numbers which belong to LinSpace are also known as Grzegorzczuk's class E_*^2 .

(b) Nondeterministic Computations

A *nondeterministic Turing machine* is a Turing machine in which the finite automaton is nondeterministic (above we have considered only deterministic machines). This means that in some states there is more than one possible action of the machine. Thus there might be many possible computations on a given input word w . A word w is *accepted* by a machine if there exists (at least one) computation leading to an accepting state. When counting the computation resources needed for a given input word (time and space) we consider the worst case. Now we can define nondeterministic counterparts of the deterministic classes. They will be denoted simply by prefixing N . For example:

$$NP = \bigcup_{c \in \mathbb{N}} N\text{Time}(n^c) \quad (\text{Nondeterministic Polynomial Time}).$$

Nondeterminism adds much power to the computations. We can *guess* a solution to a problem and then *check* its correctness. This suggests the following, more logical, definition of NP :

$$L \in NP \equiv (\exists R(x, y) \in P)(\exists p\text{-polynomial}) L = \{x \mid (\exists y, |y| \leq p(|x|)) R(x, y)\},$$

see Theorem 2.11 below.

As an example consider the complexity of the set of composite numbers. We can guess a proper divisor of a number k and check that it really is a

divisor of k . This proves that this set is in NP ; it is an open problem whether it is in P .

The most important problem in complexity theory (and one of the most important problems in contemporary mathematics) is the question whether $P = NP$. It is a part of a more general problem about the relation of nondeterministic classes to deterministic ones. We have the following trivial inclusions

$$Time(f(n)) \subseteq NTime(f(n)) \subseteq Space(f(n)),$$

but for converse simulations we have only exponential bounds. For space classes the situation is a little different:

2.5 Theorem. $NSpace(f(n)) \subseteq Space(f(n)^2)$, provided $f(n)$ can be computed in space $f(n)$. Hence $NPSpace = PSpace$.

The next natural step is to consider the role of the universal quantifier. Traditionally this is done by taking complements of the nondeterministic classes. Thus we introduce another prefix for complexity classes, co , to denote the class of the complements of the sets in a given class. (It is superfluous in connection with deterministic classes, since they are closed under complements.) The problem whether $NP = coNP$ arises quite often in connection with problems in logic. It seems likely that the time classes are *not* closed under complements. On the other hand, quite surprisingly, a proof has been found recently showing that nondeterministic space classes *are* closed under complements.

2.6 Theorem. $coNSpace(f(n)) = NSpace(f(n))$, provided $f(n) \geq \log_2 n$ can be computed in nondeterministic space $f(n)$.

However, it is still open whether $Space(f(n)) = NSpace(f(n))$, in particular whether $LogSpace = NLogSpace$.

(c) Degrees and NP -completeness

A set (language) A is *polynomially reducible* to a set B if there exists a polynomial time computable function F (i.e. $F \in \square_1^P$) such that $F^{-1}(B) = A$. Clearly, this relation is a quasiordering, hence defines a structure of degrees. The smallest degree is P .

2.7 Theorem. There exists the largest degree contained in NP .

The sets in this degree are called *NP -complete*. Not only do NP -complete sets exist, but they occur very naturally. There are hundreds of naturally defined NP -complete problems. This fact is important for the $P \stackrel{?}{=} NP$ problem:

we know that NP -complete sets are *not* in P , if $P \neq NP$ (which is accepted by most people almost as an axiom). Another consequence is that the statement " $P \neq NP$ " is equivalent to a Π_2 formula.

Examples of NP -complete Problems.

(1) **CLIQUE** – the set of pairs (G, k) where G is a graph containing a clique (i.e. a complete subgraph) of size k . (We assume that graphs are suitably coded as strings in a finite alphabet.) This is a typical NP -complete set, since it is connected with an optimization problem: find the largest clique in a given graph. We shall talk about optimization problems at the end of Sect. 4.

(2) **HAMILTONIAN GRAPHS** – the set of graphs that contain a cycle containing all vertices.

(3) The set of triples (a, b, c) of integers $a, b, c > 0$ such that the equation

$$(c.1) \quad ax^2 + by = c$$

is solvable in natural numbers.

The last example shows that there are NP -complete problems expressible by bounded formulae, since, clearly, we can bound any solution x, y to (c.1) by c . This sets some limitations on the provability of Matiyasevič's theorem in weak fragments. Consider the formula

$$\varphi(a, b, c) \equiv_{df} (\forall x, y \leq c)(ax^2 + by \neq c).$$

By Matiyasevič's theorem there exists an open formula $\psi(a, b, c, \mathbf{z})$ such that

$$N \models \varphi(a, b, c) \equiv (\exists \mathbf{z})\psi(a, b, c, \mathbf{z}),$$

where $\mathbf{z} = z_1, \dots, z_k$. Suppose this equivalence was provable in $I\Sigma_0 + \Omega_1$. Then in particular,

$$I\Sigma_0 + \Omega_1 \vdash \varphi(a, b, c) \rightarrow (\exists \mathbf{z})\psi(a, b, c, \mathbf{z}).$$

Hence, by Theorem 1.4, we can add bounds to the variables $\mathbf{z}$ with the following form

$$|z_1|, \dots, |z_k| \leq p(|a|, |b|, |c|),$$

where p is a polynomial. Thus $\varphi(a, b, c)$ would be equivalent to

$$(\exists \mathbf{z}, |z_1|, \dots, |z_k| \leq p(|a|, |b|, |c|))\psi(a, b, c, \mathbf{z}),$$

which would prove that $\varphi(a, b, c)$ defines a set in NP . It can easily be proved that if the complement of an NP -complete set is in NP then $NP = coNP$.

Hence if $NP \neq coNP$, then the bounded formula $\varphi(a, b, c)$ is not equivalent to an existential formula in $I\Sigma_0 + \Omega_1$.

Let us mention only one more result about the degrees.

2.8 Theorem. If $P \neq NP$, then the degrees below NP are dense; in particular, there is a degree strictly between P and NP -complete sets.

(d) Oracle Computations

An *oracle Turing machine* is a Turing machine with a special tape, called an *oracle tape*, and special *oracle states*. In an oracle state the machine can perform exactly two actions. Given a set A , an *oracle*, the machine computes in the usual manner except when it reaches an oracle state. Then it asks whether the current word on the oracle tape belongs to A and acts according to the answer. Thus if the non-oracle states are deterministic, the computation is deterministic (provided A is fixed). In order to limit the amount of information provided by an oracle in the case of linear time computations, we shall assume that before the next query is asked, the previous one is erased.

Let C be a complexity class and let A be a set (an oracle). We denote by $C(A)$ the class defined in the same way as C except that the Turing machines are augmented with the oracle A . We say that $C(A)$ is C *relativized to A* . We can also take a class A of oracles, in particular a complexity class, and consider

$$C(A) =_{df} \bigcup_{A \in A} C(A).$$

Almost all statements about relations between complexity classes relativize, which means that they remain true if we add an oracle. On the other hand there is the following result which shows that the $P \stackrel{?}{=} NP$ problem cannot be solved without going deep into the combinatorial structure of polynomial time computations.

2.9 Theorem. There are oracles A and B such that $P(A) = NP(A)$ and $P(B) \neq NP(B)$.

Similar “independence” can be shown for other open problems about complexity classes.

(e) The Linear Time Hierarchy and the Polynomial Hierarchy

2.10 Definition.

$$\Sigma_0^{lin} =_{df} LinTime;$$

$$\Sigma_{i+1}^{lin} =_{df} NLinTime(\Sigma_i^{lin});$$

$$LinH =_{df} \bigcup_i \Sigma_i^{lin} \quad (Linear\ Time\ Hierarchy);$$

$$\Sigma_0^p =_{df} P;$$

$$\Delta_{i+1}^p =_{df} P(\Sigma_i^p);$$

$$\Sigma_{i+1}^p =_{df} NP(\Sigma_i^p);$$

$$PH =_{df} \bigcup_i \Sigma_i^p \quad (Polynomial\ (Time)\ Hierarchy);$$

$$\square_1^p - \text{functions computable in polynomial time;}$$

$$\square_{i+1}^p =_{df} \square_1^p(\Sigma_i^p) - \text{functions computable in polynomial time with oracles from } \Sigma_i^p.$$

In particular, we have $\Sigma_1^{lin} = NLinTime$ and $\Sigma_1^p = NP$. As usual we write Π_i^p for $co\Sigma_i^p$ and similarly for the classes of the Linear Time Hierarchy. We have also the same diagram of inclusions as for the Arithmetic Hierarchy, but it is an open problem whether any of these inclusions is strict, both for the Polynomial Hierarchy and for the Linear Time Hierarchy.

We shall show that the “logical” definition of NP extends to all classes of the Polynomial Hierarchy and that there is a similar representation for the classes of the Linear Time Hierarchy. In order to simplify the formulae, let us write

$$(\exists^{lin} x)\varphi \text{ for } (\exists x, |x| \leq \ell(|y_1|, \dots, |y_i|))\varphi,$$

where $\ell(|y_1|, \dots, |y_i|)$ is a linear function (with positive coefficients) and $y_1, \dots, y_i$ are the free variables of φ . $\forall^{lin}$ has the dual meaning; $\exists^p$ and $\forall^p$ are defined in the same way except that we use polynomials instead of linear functions.

2.11 Theorem. (a) Σ_i^{lin} is the class of sets defined by formulae of the form

$$(\exists^{lin} x_1)(\forall^{lin} x_2) \dots (Q^{lin} x_i) R(w, x_1, x_2, \dots, x_i),$$

where R is in $LinTime$, and Q is $\exists$ if i is odd and $\forall$ if i is even.

(b) Σ_i^p is the class of sets defined by formulae of the form

$$(\exists^p x_1)(\forall^p x_2) \dots (Q^p x_i) R(w, x_1, x_2, \dots, x_i),$$

where R is in P and Q is $\exists$ if i is odd and $\forall$ if i is even.

Proof. We shall sketch a proof of (a), since (b) is similar and, in fact, simpler. For $i = 0$ it is trivial. Consider $i = 1$. Thus we have a nondeterministic Turing

machine M running in linear time and an oracle for a set in $LinTime$. We would like to express the set accepted by the oracle machine by a formula $(\exists^{lin} x_1)R(w, x_1)$ with R in $LinTime$. We shall use x_1 to code the following information:

- (1) the nondeterministic decisions of the machine M ,
- (2) the queries asked by the machine, and
- (3) the answers of the oracle.

R will be computed by a machine which simulates M *deterministically* using x_1 and using the subroutine for the oracle. Note that the total time needed to run the subroutines is linear on the input size, since the sum of the lengths of the oracle queries is linear.

Consider a general i . We use x_1 to code the same information. By the induction assumption, the oracle answers are determined by a formula

$$(e.1) \quad (\exists^{lin} x'_1)(\forall^{lin} x'_2) \dots (Q^{lin} x'_{i-1})R'(w', x'_1, x'_2, \dots, x'_{i-1}).$$

We need to express that the oracle answers coded by x_1 are correct. This can be done by a conjunction of instances of (e.1) and its negation where w' runs through the oracle answers. The negations of (e.1) are equivalent to formulae starting with a universal quantifier, hence their conjunction can be easily expressed as a single formula of the form

$$(\forall^{lin} x''_1)(\exists^{lin} x''_2) \dots (Q^{lin} x''_{i-1})R''(w'', x''_1, x''_2, \dots, x''_{i-1}).$$

The nonnegated instances of the formula (e.1) can also be expressed by a single formula. Again the trick is that the sum of the lengths of the oracle queries is linear, hence also the sum of the lengths of the variables x'_1 can be bounded using a linear function. We omit the details. The converse implication is easy. $\square$

We shall show that $LinH$ consists of sets definable by Σ_0 formulae (in the language L_0 ; we consider only subsets of natural numbers). Later, in Sect. 4, we shall show that PH consists of sets definable by bounded formulae in L_2 . That is why these complexity classes are so important for us. We shall denote by Σ_0^N the subsets of natural numbers definable by Σ_0 formulae.

2.12 Lemma. $LinTime \subseteq \Sigma_0^N$.

Proof. We shall assume that a suitable coding of sequences can be expressed using Σ_0 formulae. This will be done with details in Sect. 3. (Note that our task in Sect. 3 will be harder, since we shall develop coding of sequences in the *theory* $I\Sigma_0$.) Furthermore we need the function $nuon$ which counts the number of ones in a binary expansion of a number. A Σ_0 formula defining this function is also constructed in Sect. 3.

Our goal is to express, by a Σ_0 formula $\varphi(w)$, that a machine M accepts a word w provided that M runs in linear time. To this end we need to code the computation of M by a sequence s whose length is linear in the length of w . This is because the polynomial bound on the size of numbers, as it is in Σ_0 , imposes a linear bound on their length and *vice versa*. We code sequences essentially by bits in binary expansions (see Sect. 3 for details).

Suppose M has k heads. The sequence s will consist of $(k+1)$ -tuples, one for each computation step. The first member of the $(k+1)$ -tuple will code the state of M . The remaining members will be pairs, one for each head. They will code the currently read symbol and the direction in which the head will move. Clearly the length of (the code of) s is linear in the length of w . The sequence s does not contain information about the positions of heads on tapes explicitly, but this can be defined from s by a Σ_0 formula. The position of a head in a computation step is the number of its moves to the right minus the number of its moves to the left prior to this step. The number of moves to the right (similarly to the left) is the number of occurrences of the symbol coding these moves in s . This can be easily reduced to the function *nuon*. Now for a given head and a given computation step (as coded in s), we can determine using Σ_0 formulae whether the currently read cell has been visited before, and if it has, then we can determine the last visit.

In order to state that s is an accepting computation, we have to impose conditions on $(k+2)$ -tuples

$$c, c^+, a, c_2, \dots, c_k,$$

where c is an element of s , c^+ is its successor, a is the symbol in the input word w read in the step corresponding to c and $c_2, \dots, c_k$ are elements of s which correspond to the last visits of work heads of the cells read in c , (some of them may be missing if some heads visit some cells for the first time). Since the automaton of M is finite and all the alphabets are finite, there are finitely many possibilities for such a $(k+2)$ -tuple and they can be listed. Finally, of course, we add that the last state is an accepting state. $\square$

2.13 Lemma. $\text{LinH} \subseteq \Sigma_0^N$.

Proof. This follows from Lemma 2.12, Theorem 2.11 and the observation made above that the growth rate of bounds in Theorem 2.11 and those of Σ_0 are the same. $\square$

Before showing the converse inclusion, we prove an important result.

(f) Nepomnjaščij's Theorem

We need to introduce another type of complexity class, classes defined by simultaneous resource bounds:

$TimeSpace(f(n), g(n))$ is the class of sets which can be accepted by Turing machines running *simultaneously* in time $f(n)$ and space $g(n)$. Note that this class is in general smaller than $Time(f(n)) \cap Space(g(n))$.

2.14 Theorem. Let ε be a rational number, $0 < \varepsilon < 1$, and let a be a positive integer. Then $TimeSpace(n^a, n^\varepsilon) \subseteq LinH$.

Proof. First we shall sketch an idea of the proof in which we shall assume that the numbers n^α needed in the proof, with α rational, are integers and that they are computable in time n^α (which is not always true).

Let ε , $0 < \varepsilon < 1$, be given. We shall prove by induction on k that $TimeSpace(n^{k(1-\varepsilon)}, n^\varepsilon) \subseteq LinH$. For $k = 1$ it is trivial. Suppose it holds for k . Let M be a Turing machine. We shall call a *configuration* the complete description of the machine state in a computation step, i.e. the state of the finite control, the position of heads, the content of the work tapes; only the input word is not included. We assume that configurations are naturally encoded as strings over a finite alphabet. Suppose M runs in space n^ε . Let an input word w and two configurations c_1, c_2 of M be given. Then we can test in $TimeSpace(n^{k(1-\varepsilon)}, n^\varepsilon)$ whether M will reach c_2 from c_1 in $n^{k(1-\varepsilon)}$ steps: just simulate M . Hence, by the induction hypothesis, this condition on w, c_1, c_2 can be tested in $LinH$. Now let M' be a machine running in time $n^{(k+1)(1-\varepsilon)}$ and space n^ε . To show that M' accepts a set in $LinH$ we shall simulate M' by an oracle nondeterministic Turing machine K running in linear time. Let w be an input word. First, K constructs nondeterministically $n^{1-\varepsilon} + 1$ strings. Then it checks that:

- (1) the first string is the initial configuration of M' ;
- (2) every pair of consecutive strings are configurations of M' such that M' can reach the second one from the first one in $n^{k(1-\varepsilon)}$ steps;
- (3) the final string is an accepting configuration.

Here (1) and (3) can be done deterministically in linear time (always assuming a reasonable coding of configurations). For (2) it will use the oracle shown to exist above. We need only to show that K runs in linear time. The key idea of the whole proof is that the total length of the $n^{1-\varepsilon} + 1$ configurations is linear in n . This is because M' runs in space n^ε ; hence the strings representing configurations of M' have lengths linear in n^ε .

We have omitted several details, but we shall only comment on the one mentioned above. The problem is how to compute $\lceil n^{k(1-\varepsilon)} \rceil$. The point is that, due to nondeterminism, we do not have to count the number of steps during the simulation at all. More precisely, in the inductive construction we need only a relation $R(w, c_1, c_2)$ in $LinH$ such that

- (1) if M will reach c_2 from c_1 in $n^{k(1-\varepsilon)}$ steps on input w then $R(w, c_1, c_2)$, and

- (2) if $R(w, c_1, c_2)$ then M will reach c_2 from c_1 on input w (using possibly even more than $n^{k(1-\epsilon)}$ steps).

The simulating machine can guess the intermediate configurations that are evenly (as much as possible) distributed. $\square$

We shall use this theorem directly only in a very special case (the next corollary). However the *idea* of using quantifiers in order to code longer than linear computations is very useful. We shall use it in Sect. 3 to construct a Σ_0 definition of the function *nuon*.

2.15 Corollary. $\text{LogSpace} \subseteq \text{LinH}$.

Proof. We know that $\text{LogSpace} \subseteq P$, but the proof of this fact gives more:

$$\text{LogSpace} \subseteq \bigcup_k \text{TimeSpace}(n^k, \log n).$$

Hence, by Theorem 2.14, $\text{LogSpace} \subseteq \text{LinH}$. $\square$

2.16 Theorem. Σ_0 definable sets are just the sets in the Linear Time Hierarchy, i.e. $\Sigma_0^N = \text{LinH}$.

Proof. By Lemma 2.12, it remains to prove $\Sigma_0^N \subseteq \text{LinH}$. By Theorem 2.11 and since the bounded quantification is in both cases of the same growth rate, we only need to show that the open arithmetical formulae define sets in LinH . It can be easily shown that LinH is closed under set operations, hence the question reduces to atomic formulae. The relations $=$ and $\leq$ are easily computable in linear time and so is addition. We do not know whether multiplication of two numbers is computable in linear time (probably it is not), but we can show that it is computable in logarithmic space. This can be proved by analyzing the school algorithm for multiplication. Suppose we multiply a number of length n with a number of length m in binary. First we produce m numbers of length at most $n + m$ and then we add them. This needs space of the order mn . However to add these numbers we do not have to write them down. We only have to remember the current row and the column and we can always compute the entry at this place. To remember the position, we need about $\log m + \log n$ bits. Furthermore, we have to remember the carry, which is of length at most $\log m$. Thus the total space needed to compute the product is logarithmic in the length of the numbers. It is also easily seen that we do not need to compute the multiplication as an operation; we only need to compute its graph. Hence, by Corollary 2.15 the graph of multiplication is in LinH . $\square$

(g) The Diagonal Method for Separating Complexity Classes

The most important problems in complexity theory can be stated as problems about separating some complexity classes (which means showing that they are different). In Bounded Arithmetic the most important problems are also about separating fragments of Bounded Arithmetic. There are two kinds of such problems. The first kind concerns fragments defined by restricting the quantifier complexity of an induction schema. For instance we do not know whether $IE_i = IE_{i+1}$, $S_2^i = S_2^{i+1}$ and $T_2^i = T_2^{i+1}$. In the problems of the second kind, the meaning of "separation" is Π_1 nonconservation. The typical problem here is whether $I\Sigma_0 + \Omega_1$ is Π_1 conservative over $I\Sigma_0$. For each fragment of Bounded Arithmetic there is a naturally assigned complexity class: the class of sets which are definable by formulae for which the induction schema is provable in the fragment. Thus Σ_0^N , i.e. $LinH$, corresponds to $I\Sigma_0$, PH corresponds to $I\Sigma_0 + \Omega_1$, etc. The separation problems for fragments and those for the corresponding complexity classes seem to be connected, though we know very little about it (a connection is shown at the end of Sect. 4). (Let us note that one can consider different schemas of induction and different base theories, thus the correspondence between fragments and complexity classes is not one-to-one.)

Here we would like to discuss a method on which almost all separation results in complexity rely. It is called *diagonalization*. Its applicability is limited; a rule of thumb is that we can separate only complexity classes defined using bounds to the same resources. The method turns out to be useless for time-space and determinism-nondeterminism problems. A natural question to ask now is what corresponds to this method in fragments of arithmetic. In metamathematics of arithmetic diagonalization is well-known: it is the essence of Gödel's theorems. It seems that it has similar limitations in fragments of Bounded Arithmetic. In particular, no fragments have been separated for which the corresponding problem in complexity theory is open.

We need some technical concepts before stating the basic results which use diagonalization. A function f is *space constructible* if there exists a Turing machine with one work tape which on each standard input of length n (say a word consisting of n 0's) uses a space of exactly $f(n)$. A function f is *time constructible* if there is a Turing machine which uses a time of exactly $f(n)$ on each standard input of length n . These concepts are needed to avoid having possible pathological functions as space or time bounds; the usual bounding functions are easily shown to be space and time constructible.

2.17 Theorem. (i) Let f and g be space constructible, $f(n) \geq \log_2 n$ and suppose

$$\liminf f(n)/g(n) = 0.$$

Then

$$Space(g(n)) \setminus Space(f(n)) \neq \emptyset.$$

(ii) Let f and g be time constructible and suppose

$$\liminf f(n)/g(n) = 0.$$

Then

$$\text{Time}(g(n) \log_2 g(n)) \setminus \text{Time}(f(n)) \neq \emptyset.$$

Thus we have $\text{LinTime} \subsetneq P \subsetneq E \subsetneq \text{Exp}$ and $\text{LogSpace} \subsetneq \text{LinSpace} \subsetneq \text{PSpace}$. An interesting consequence of $\text{LogSpace} \subsetneq \text{PSpace}$ is that at least one of the inclusions in the following sequence is strict:

$$\text{LogSpace} \subseteq P \subseteq NP \subseteq \text{PSpace}.$$

But for each of them it is an open problem.

The idea of the **proofs** of both statements is essentially the same. Take a suitable coding of Turing machines as strings in an alphabet Σ . Let $\#$ be a symbol not in Σ . Construct a Turing machine M which works as follows. On each input it will simulate a machine computing the bound $g(n)$ and stop if the bound is achieved, so M is space (resp. time) bounded by g . If an input w is a code of a machine K followed by a string of $\#$'s, at the same time (say, on an extra tape) M will simulate K on w in space (resp. time) limit $g(|w|)$. If the simulation of K is finished before M runs out of space (resp. time), then M accepts w iff K does not. This situation must happen for each K which has a space (resp. time) bound $f(n)$, since the bound $g(n)$ for M is larger for a sufficiently long w . Thus the language accepted by M is different from the language accepted by each such K . The reason for $g(n) \log_2 g(n)$ instead of just $g(n)$ in (ii) is that we have to simulate by M machines with an arbitrary number of tapes and this is not possible without an increase in time.

The proof of the corresponding results for nondeterministic classes is more difficult. The problem is how to arrange that M accepts the input iff K does not. In the case of nondeterministic space classes we could use the fact that they are closed under complements (Theorem 2.6), but this cannot be used in the case of nondeterministic time classes. There is, however, a nice trick for avoiding this problem. This trick works even in more general situations.

2.18 Theorem. Let C be a class of languages in an alphabet Σ , with $0 \in \Sigma$, let L be a language in the alphabet Σ , let r be a mapping assigning a word in the alphabet Σ to each language in C and let z be a mapping from C to natural numbers. Suppose that for every $K \in C$

- (a) $r(K) \cap 0^{z(K)} \in L \equiv r(K) \notin K$,
- (b) $(\forall j, 0 \leq j < z(K))(r(K) \cap 0^j \in L \equiv r(K) \cap 0^{j+1} \in K)$.

Then $L \notin C$. (0^j denotes the string of j 0's.)

Proof. Suppose $L = K$ for some $K \in C$. Then

$$r(K) \frown 0^{j+1} \in L \equiv r(K) \frown 0^{j+1} \in K,$$

hence by (b),

$$r(K) \in L \equiv r(K) \frown 0^{z(K)} \in L.$$

Thus by (a)

$$r(K) \in L \equiv r(K) \notin K,$$

which is a contradiction. $\square$

In the usual diagonalization we have only the simple condition $r(K) \in L \equiv r(K) \notin K$, which is, however, sometimes more difficult to satisfy than (a) and (b) above. This is because we can take z to be an exponentially growing function. Then in order to determine whether $r(K) \frown 0^{z(K)} \in L$, we need to decide whether a much smaller word $r(K)$ belongs to K . Theorem 2.18 can be used not only for nondeterministic classes, but also for classes with a larger quantifier prefix. We shall show only a very special case of such a theorem.

2.19 Theorem. For every $i \geq 0$, $\Sigma_i^{lin} \subsetneq \Sigma_i^p$.

Proof. Let i be given. We shall apply Theorem 2.18 with $C = \Sigma_i^{lin}$. We shall also use the representation of Σ_i^{lin} and Σ_i^p given by Theorem 2.11. Thus every $K \in \Sigma_i^{lin}$ is defined by a formula

$$(g.1) \quad (\exists^{lin} x_1)(\forall^{lin} x_2) \dots (Q^{lin} x_i) R(w, x_1, x_2, \dots, x_i),$$

where R is in *LinTime*. The diagonal language L will be defined by a formula of the form

$$(\exists x_1, |x_1| \leq |w|^2)(\forall x_2, |x_2| \leq |w|^2) \dots (Q x_i, |x_i| \leq |w|^2) S(w, x_1, x_2, \dots, x_i),$$

with $S(w, x_1, x_2, \dots, x_i)$ computable in time $|w|^4$, for $|x_1|, \dots, |x_i| \leq |w|^2$, by a two tape deterministic Turing machine. We shall use the result that an $f(n)$ time bounded multitape Turing machine can be simulated by a $cf(n) \log_2 f(n)$ time-bounded two tape Turing machine, with c a constant. (We do not need such a good simulation; a simple polynomial simulation is sufficient, but then $|w|^4$ must be replaced by a polynomial of higher degree.) Thus in order to simulate a linear time Turing machine, we need asymptotically less than n^2 time.

Let $K \in \Sigma_i^{lin}$ be defined by a formula (g.1) with $\ell_1(|w|), \dots, \ell_i(|w|)$ as the implicit linear bounds for the quantifiers. We take $r(K)$ to be a word which contains a code of the machine computing R and the codes of the

linear bounds, and moreover assume it is sufficiently long. Namely we want, for $n \geq |r(K)|$,

$$\ell_1(n+1) \leq n^2, \dots, \ell_i(n+1) \leq n^2,$$

and that a computation of the predicate

$$|x_1| \leq \ell_1(|w|) \& ((|x_2| \leq \ell_2(|w|) \rightarrow (|x_3| \leq \ell_3(|w|) \& (\dots \\ \dots R(w, x_1, x_2, \dots, x_i) \dots)),$$

for $|x_1|, \dots, |x_i| \leq |w|^2$, $|w| \geq n$ be possible in time $(|w| - 1)^4$. Furthermore we construct $r(K)$ so that it does not contain occurrences of 0. Clearly each such a K can be computed in exponential time. There is a fixed exponential bound $f(n)$ such that each such K can be computed within this bound by a two tape Turing machine for *sufficiently long inputs*. Again, sufficiently long means longer than $|r(K)|$. We define $z(K) = f(|r(K)|)$.

Now we define the machine M which determines the predicate S . Given an input $(w \smallfrown 0^j, x_1, x_2, \dots, x_i)$, where the last symbol of w is not 0, M will compare j and $f(|w|)$. If $j > f(|w|)$ then what it does is not important. If $j = f(|w|)$, then M will try to interpret w as a code of a machine and linear bounds which determine a language K defined by a formula of the form (g.1). Again it is important only to consider the case where w is such a code. Then M will (deterministically) decide whether $w \in K$. This is possible in time $|w \smallfrown 0^j|^2$, since $j = f(|w|)$. M accepts such an input iff $w \notin K$. (Thus in this case $x_1, \dots, x_i$ are not used at all.) Clearly this ensures condition (a) of Theorem 2.18. Now suppose that $j < f(|w|)$. In this case M will also try to interpret w as a code of a machine and linear bounds, but now it will compute the predicate

$$|x_1| \leq \ell_1(|u|) \& ((|x_2| \leq \ell_2(|u|) \rightarrow (|x_3| \leq \ell_3(|u|) \& (\dots \\ \dots R(u, x_1, x_2, \dots, x_i) \dots)),$$

for $u = w \smallfrown 0^{j+1}$. This can be done in time $|w \smallfrown 0^j|^4$, if $w = r(K)$ (where K is the language determined by the code w) and $|x_1|, \dots, |x_i| \leq |w \smallfrown 0^j|^2$. Hence we have

$$(\exists^{\text{lin}} x_1)(\forall^{\text{lin}} x_2) \dots (Q^{\text{lin}} x_i) R(w \smallfrown 0^{j+1}, x_1, x_2, \dots, x_i),$$

if and only if

$$(\exists x_1, |x_1| \leq |w \smallfrown 0^j|^2)(\forall x_2, |x_2| \leq |w \smallfrown 0^j|^2) \dots (Q x_i, |x_i| \leq |w \smallfrown 0^j|^2) \\ S(w \smallfrown 0^j, x_1, x_2, \dots, x_i).$$

This proves that condition (b) of Theorem 2.18 is satisfied. Thus L is a language which belongs to PH but not to $LinH$. $\square$

We say that $LinH$ (PH resp.) collapses if for some i , $LinH = \Sigma_i^{lin}$ ($PH = \Sigma_i^p$ resp.). This is equivalent to $\Sigma_i^{lin} = \Pi_i^{lin}$ ($\Sigma_i^p = \Pi_i^p$ resp.). We conjecture that neither $LinH$ nor PH collapses and that $LinH$ is a proper subclass of PH . But we have only the following result.

2.20 Corollary. Either $LinH \subsetneq PH$ or $LinH$ does not collapse.

Proof. Suppose $LinH = \Sigma_i^{lin}$. Then $LinH \subsetneq PH$, since, by Theorem 2.19, $\Sigma_i^{lin} \subsetneq \Sigma_i^p \subseteq PH$. $\square$

Let us define Σ_i^f , for f a function, to be the class of sets definable by formulae of the form

$$(\exists x_1, |x_1| \leq cf(|w|))(\forall x_2, |x_2| \leq cf(|w|)) \dots (Qx_i, |x_i| \leq cf(|w|)) \\ R(w, x_1, x_2, \dots, x_i),$$

where c is a constant, $R(w, x_1, x_2, \dots, x_i)$ is computable in time $cf(|w|)$ for every $x_1, \dots, x_i$, $|x_1| \leq cf(|w|), \dots, |x_i| \leq cf(|w|)$ and where Q is $\exists$ if i is odd and $\forall$ if i is even. Let

$$TH(f(n)) =_{df} \bigcup_i \Sigma_i^f \quad (\text{Time Hierarchy } f).$$

The proof of Theorem 2.19 can be generalized to prove a statement similar to Theorem 2.17 for classes Σ_i^f and Σ_i^g where g grows faster than f for each fixed i . For classes $TH(f(n))$ this is an open problem (as we have mentioned above we do not know whether $LinH = PH$). Our last theorem shows that we can separate PH from slightly subexponential hierarchies. The bound 2^{n^ϵ} can be further decreased by iterating our method.

2.21 Theorem. For $\epsilon > 0$, $PH \subsetneq TH(2^{n^\epsilon})$.

First we shall prove some lemmas. Let L be a language and suppose that 0 is not in the alphabet of L ; let f be a function such that $f(n) \geq n$ for all n . We define

$$L[f] =_{df} \{w \frown 0^{f(|w|)-|w|} \mid w \in L\}.$$

2.22 Lemma. Let f, g be time constructible functions.

- (a) If $f(n) \geq g(n)$ for $n > n_0$, then for every $i \geq 0$, $\Sigma_i^g \subseteq \Sigma_i^f$.
- (b) For every $i \geq 0$, $L \in \Sigma_i^f$ iff $L[f] \in \Sigma_i^{lin}$.

Proof. (a) is trivial, (b) follows easily from the definition of Σ_i^f and from the characterization of Σ_i^{lin} (Theorem 2.11 (a)). $\square$

2.23 Lemma. Let f, g be time constructible functions; let g be growing faster than every polynomial. Suppose $TH(gf(n)) = PH$. Then $TH(ff(n)) \subseteq PH$.

Proof. Let the assumption of the lemma be satisfied. Let $L \in TH(ff(n))$. Using Lemma 2.22(b) we get first $L[ff(n)] \in LinH$ and then $L[f] \in TH(f)$. Since $TH(gf(n)) = PH$, we get $L[f] \in PH$. Using Lemma 2.22(b) again we have $L[pf(n)] \in LinH$ for some polynomial p . Hence $L \in TH(pf(n))$. Since g grows faster than any polynomial we have

$$TH(pf(n)) \subseteq TH(gf(n)) = PH.$$

Hence $L \in PH$. □

Proof of Theorem 22.1. Let $\varepsilon > 0$ be given. Let

$$\begin{aligned} f(n) &= 2^{\lfloor n^{\varepsilon/2} \rfloor}, \\ g(n) &= 2^{\lfloor \log_2 n \rfloor^2}. \end{aligned}$$

Suppose $PH = TH(2^{\lfloor n^{\varepsilon} \rfloor})$. Then we have

$$PH \subseteq TH(gf(n)) \subseteq TH(2^{\lfloor n^{\varepsilon} \rfloor}) = PH,$$

since it can be easily computed that $gf(n)$ grows faster than any polynomial but is smaller than $2^{\lfloor n^{\varepsilon} \rfloor}$. Thus $TH(gf(n)) = PH$. By Lemma 2.23 we have $TH(ff(n)) \subseteq PH$. We have the following inclusions:

$$PH \subseteq Exp \subsetneq Time \left(2^{\lfloor n^{\varepsilon'} \rfloor} \right) \subseteq TH \left(2^{\lfloor n^{\varepsilon'} \rfloor} \right) \subseteq TH(ff(n)) \subseteq PH,$$

where $\varepsilon/2 > \varepsilon' > 0$. The first inclusion is a natural simulation, the second one follows from Theorem 2.17(ii), the third one is trivial and the fourth one follows from a simple computation and Lemma 2.22(a). Thus we get a contradiction. □

*

We have defined several complexity classes, only a few of which will be mentioned in the following sections. The larger variety should help the reader to understand better the environment in which we are working, if we consider weak fragments. This short survey, however, cannot replace a textbook on such a broad subject.

3. Exponentiation, Coding Sequences and Formalization of Syntax in $I\Sigma_0$

(a) Introduction

In this chapter we shall present some basic constructions available in $I\Sigma_0$. First we want to show that it is not necessary to add the relation $x^y = z$ to $I\Sigma_0$, since it is definable there. To this end we have to develop the theory of sequences a little. Once we have the exponentiation relation we will be able to define a more efficient way of coding sequences. The second important task is to describe a formalization of syntax in $I\Sigma_0$. A formalization of syntax in $I\Sigma_1$ has been presented in Chap. I, Sect. 1(b). However $I\Sigma_0$ is a much weaker theory and it is not always possible to use classical techniques here. Usually formalization is very tedious work and the reader can find very few interesting ideas there. We cannot skip this part completely however as we have to persuade the reader that such a formalization is possible in a theory which at first looks very weak. Therefore we have decided on a compromise: instead of describing a particular formalization we develop some tools which can be easily applied to most standard situations. Clearly the reader of this book is not looking for formulae, but he is interested in ideas and techniques. In this section we shall consider only combinatorial problems related to the formalization of syntax. More about the formalization will be presented in Sect. 5.

The weaker the theory the more difficult it is to prove theorems. This applies to coding and formalization as well. Many concepts can be more easily handled in $I\Sigma_0 + \Omega_1$ than in $I\Sigma_0$. Since much research is going on in $I\Sigma_0 + \Omega_1$ (or equivalent theories) it is worthwhile mentioning these simpler formalizations.

Much of our intuition about natural numbers relies on exponentiation. Thus in order to persuade the reader about the possibility of a suitable formalization of (some of) the usual reasoning on numbers in $I\Sigma_0$, we prove the properties of exponentiation in great detail. After introducing the exponentiation we shall use more sketchy arguments.

Besides exponentiation we shall define one more important Σ_0 -definable function: "the number of ones in the binary expansion". With these two concepts in hand, formalization of syntax will proceed quite smoothly.

The last subsection presents a general theorem which can be applied to formalize various syntactical concepts in $I\Sigma_0$.

In some estimates we shall use explicit numerals. The reason is that these numbers are easily computable and small.

(b) Sets and Sequences

Since $I\Sigma_0$ contains I_{open} , we know, by Chap. I, Sect. 1, that we can prove some elementary number theoretical facts. In order to be able to interpret other combinatorial objects, we need to formalize the concept of the sequence. Our task would not be so difficult, if we could start with the exponentiation relation in our language. But so far we do not know that this relation has a suitable definition in $I\Sigma_0$. An obvious requirement is that the defining formulae of the concepts be Σ_0 , otherwise we cannot use $I\Sigma_0$. Therefore we cannot use the usual definition of the exponentiation relation in $I\Sigma_1$, which is not bounded.

Our approach is similar to that of [Nelson 86]. It is based on the observation that it is possible to talk about the digits in the binary expansion of a number, though (at first) we are not able to define that the digit in question is the i -th digit. The point is that we can define that a number is a power of 2 and then we can index the digits of a number by powers of 2. Namely, the i -th digit will have index 2^i . Then, clearly, we can easily talk about the relative position of digits. To code a sequence of numbers we shall use a pair of numbers. The first number will be the number determined by the concatenation of binary expansions of the numbers to be coded. The second one will be a binary code of the markers which determine beginnings and ends of the coded numbers. In fact we code sequences of arbitrary 0–1 words in such a way, but we shall use this possibility only after we define the exponentiation relation.

Let us consider an example. Suppose we want to code a sequence of 0–1 words

0011, 101, 010.

Then we take two numbers whose binary expansion is the following

11101010,
10001001001.

The first one is the concatenation of the words above (where we have to omit the first two 0's) and the second one is a sequence of markers which determines the partition. If this pair is considered to be a code of a sequence of numbers then it will code $(3, 5, 2)$.

Now we proceed formally in $I\Sigma_0$.

3.1 Definition (x is a power of 2).

$$\text{Pow}(x) \equiv (\forall p \leq x)(\text{Prime}(p) \ \& \ p \mid x \rightarrow p = \bar{2}) \ \& \ x \geq \bar{1}.$$

The relations *being a prime* and *p divides x* used in this definition are easily definable by Σ_0 formulae. The following is also easy to prove.

3.2 Lemma ($I\Sigma_0$).

- (i) $Pow(\bar{1})$;
- (ii) $x > \bar{1} \rightarrow (Pow(x) \equiv (\exists y)(x = \bar{2}y \ \& \ Pow(y)))$;
- (iii) $Pow(x) \ \& \ Pow(y) \rightarrow Pow(xy) \ \& \ (x \mid y \vee y \mid x)$;
- (iv) $Pow(x) \ \& \ Pow(y) \ \& \ (\bar{2}v + \bar{1})x = (\bar{2}u + \bar{1})y \rightarrow x = y$;
- (v) $Pow(x) \ \& \ Pow(y) \ \& \ xz = y \rightarrow Pow(z)$;
- (vi) $x > \bar{0} \rightarrow (\exists y)(x < y \leq \bar{2}x \ \& \ Pow(y))$.

The last lemma allows us to define:

3.3 Definition (y is the least power of 2 larger than x).

$$lpw(x) \equiv y = \min\{y; y \leq \bar{2}x + \bar{1} \ \& \ x < y \ \& \ Pow(y)\}.$$

3.4 Lemma ($I\Sigma_0$).

- (i) $lpw(\bar{0}) = \bar{1}$;
- (ii) $x > \bar{0} \rightarrow lpw(x) \leq \bar{2}x$. □

After we define the exponentiation and the length function $|x|$, it will be clear that $lpw(x) = 2^{|x|}$. Recall that we have a pairing function defined, say, by

$$(u, v) = 1/2((u + v)^2 + \bar{3}u + v),$$

see I.1.18.

3.5 Definition (p is a sequence).

$$Seq(p) \equiv (\exists u, v \leq p)(p = (u, v) \ \& \ \neg \bar{2} \mid v \ \& \ lpw(u) \leq v).$$

Here u is the sequence of 0–1 words of length at least 1 and v is the sequence of markers. At present we are not able to define the i -th element in the list, but we can easily define that a number occurs in it.

3.6 Definition.

$$\begin{aligned} x \in (u, v) \equiv & (\exists y, y' \leq v)(\exists u_1, v_1 < y)(\exists u_2, v_2 \leq v)(Pow(y) \ \& \ Pow(y') \\ & \ \& \ y' \geq \bar{2} \ \& \ x < y' \ \& \ u = u_2 y' y + x y + u_1 \\ & \ \& \ v = v_2 \bar{2} y' y + y' y + y + v_1). \end{aligned}$$

Now we shall prove only the most basic properties of this coding, since we do not need more for defining and proving properties of exponentiation.

We show that the empty set has a code and that we can always add another element (number) in the sequence.

3.7 Lemma ($I\Sigma_0$).

- (i) $(\forall x)(\text{not } x \in (\bar{0}, \bar{1}))$;
- (ii) $(\forall p, z)(\text{Seq}(p) \rightarrow (\exists q \leq \bar{9}p(z + \bar{1})^2)(\text{Seq}(q) \& (\forall x)(x \in q \equiv (x \in p \vee x = z))))$.

Proof. (i) is trivial. (ii) Let p be a sequence, $p = (u, v)$. Let $k = \max(\bar{2}, lpw(z))$. Define

$$\begin{aligned} u' &= uk + z, \\ v' &= vk + \bar{1}, \\ q &= (u', v'). \end{aligned}$$

Suppose $u > 0$. Then we have

$$\begin{aligned} u' &\leq u * \bar{2}(z + \bar{1}) + z \leq \bar{3}(z + \bar{1})u, \\ v' &\leq v * \bar{2}(z + \bar{1}) + \bar{1} \leq \bar{3}(z + \bar{1})v. \end{aligned}$$

Since the pairing function is a quadratic polynomial, we have

$$q \leq \bar{9}(z + \bar{1})^2 p.$$

We leave to the reader to prove this bound for $u = 0$. Now we are going to prove $\text{Seq}(q)$. It is clear that v' is odd. To prove $lpw(u') \leq v'$, recall that we have $z < lpw(z) \leq k$ and $u < lpw(u)$ by definition and $lpw(u) \leq v$ by the assumption that $\text{Seq}(p)$. Thus

$$u' < uk + k \leq lpw(u)k \leq vk < v'.$$

Since $Pow(lpw(u)k)$, we have

$$lpw(u') \leq lpw(u)k \leq v',$$

thus $\text{Seq}(p)$.

To show $z \in q$ put $y = 1$, $y' = k$, $v_1 = u_1 = 0$. The verification of the inequalities is straightforward then.

Now assume $x \in p$. We want to show $x \in q$. Let $y, y', u_1, v_1, u_2, v_2$ be witnesses of x being in p . Then

$$\begin{aligned} u' &= u_2 y' y k + x y k + u_1 k + z, \\ v' &= \bar{2} v_2 y' y k + y' y k + y k + v_1 k + \bar{1}. \end{aligned}$$

Thus the witnesses for x being in q are respectively

$$y k, \quad y', \quad u_1 k + z, \quad v_1 k + \bar{1}, \quad u_2, \quad v_2.$$

We only have to check that

$$u_1 k + z < yk$$

and

$$v_1 k + \bar{1} < yk.$$

The inequalities are derived as follows

$$u_1 k + z < u_1 k + k = (u_1 + \bar{1})k \leq yk,$$

and similarly the second one.

Now assume $x \in q$. We want to show $x \in p$ or $x = z$. Let us use $y, y', u_1, v_1, u_2, v_2$ for witnesses of x being in q . We consider two cases.

Case 1, $v_1 = 0$. Since v is odd and y' is even, y must be odd. As $Pow(y)$, we have $y = 1$. Thus

$$vk = v' - \bar{1} = \bar{2}v_2 y' + y' = (\bar{2}v_2 + \bar{1})y',$$

hence, since v is odd, $k = y'$. Then

$$uk + z = u' = u_2 k + x + u_1,$$

but $u_1 = 0$, since $u_1 < y = 1$, and $x < y' = k$, $z < k$. Thus $x = z$ by the uniqueness of the remainder (see I.1.15(4)).

Case 2, $v_1 > 0$. Then we have

$$vk = v' - \bar{1} = (\bar{2}v_2 y' + y' + \bar{1})y + v_1 - \bar{1}.$$

Since $v_1 - \bar{1} < y$, it is not possible that $y \mid k$. As y is a power of 2, we have $k \mid y$, (by Lemma 3.2(iii)). Hence $k \mid (v_1 - \bar{1})$. Thus we have

$$v = \bar{2}v_2 y' y k^{-1} + y' y k^{-1} + (v_1 - \bar{1})k^{-1},$$

which is not a formula in the language of arithmetic, but clearly can be expressed by such a formula. Similarly we have

$$uk = u_2 y' y + xy + u_1 - z.$$

Since $z < k$ and $k \mid y$, it must be $z \leq u_1$. Thus we can write

$$u = u_2 y' y k^{-1} + xy k^{-1} + (u_1 - z)k^{-1}.$$

All the inequalities for new witnesses follow immediately from the inequalities for the original witnesses. Thus $x \in p$. $\square$

(c) The Exponentiation Relation

Our aim is to construct a Σ_0 formula $Exp(x, y, z)$ and to prove in $I\Sigma_0$ the following formulae:

$$(c.1) \quad Exp(x, \bar{0}, z) \equiv z = \bar{1};$$

$$(c.2) \quad Exp(x, y + \bar{1}, z) \equiv (\exists v)(Exp(x, y, v) \& z = vx).$$

Before starting the construction of the formula, let us observe that (c.1) and (c.2) imply some natural properties of Exp .

3.8 Lemma ($I\Sigma_0$).

- (i) $Exp(\bar{m}, \bar{n}, \bar{k}) \equiv \overline{m^n} = \bar{k}$;
- (ii) $Exp(x, y, z) \& Exp(x, y, z') \rightarrow z = z'$;
- (iii) $Exp(x, y, z) \& v \leq y \rightarrow (\exists w \leq z) Exp(x, v, w)$.

Proof. (i) – follows from (c.1), (c.2) and a similar fact for multiplication.

(ii) We would like to prove it by induction over y . We can use x as a parameter, but what to do with z and z' ? The way out is to take the following formula instead of (ii):

$$(\forall z, z' < v)(Exp(x, y, z) \& Exp(x, y, z') \rightarrow z = z'),$$

and we can use v as a parameter of the induction.

(iii) is quite similar. □

The problem with the natural definition which uses recursion is that the size of the sequence that codes the course of values cannot be bounded by a polynomial, which is a necessary requirement for a Σ_0 formula. To overcome this difficulty we shall use sequences in which elements grow faster and hence they are shorter. The trick is to use the recursion on (binary) notation instead of the ordinary recursion.

Example. Let $y = 1101$ in binary; then we compute

$$x^0 = 1, \quad x^1 = x, \quad x^{11} = x^2 * x, \quad x^{110} = (x^2 * x)^2, \quad x^{1101} = ((x^2 * x)^2)^2 * x.$$

First we define an auxiliary formula (this is similar but not the same formula as in I.1.49).

3.9 Definition.

$$\begin{aligned} Exseq(x, s) \equiv & Seq(s) \& (\bar{1}, \bar{0}) \in s \& (\forall y, z \leq s)[(z, y) \in s \rightarrow \\ & \rightarrow (z = \bar{1} \& y = \bar{0}) \vee (y > \bar{0} \& (\exists v, w \leq s)((y = \bar{2}v \& z = w^2 \\ & \& (w, v) \in s) \vee (y = \bar{2}v + \bar{1} \& z = w^2 x \& (w, v) \in s)))]]. \end{aligned}$$

3.10 Lemma ($I\Sigma_0$).

$$x \geq \bar{2} \ \& \ Exseq(x, s) \ \& \ (z, y) \in s \rightarrow (z, y) \leq \bar{4}z^2.$$

Proof. Using induction on y we get $y \leq z$ and $z \geq 1$. Then an easy computation gives

$$(z, y) \leq (z, z) \leq \bar{4}z^2. \quad \square$$

3.11 Lemma ($I\Sigma_0$).

$$(\forall x, s)(\forall z, y \leq s)(\exists t \leq \bar{30}z^{16})(x \geq \bar{2} \ \& \ Exseq(x, s) \ \& \ (z, y) \in s \rightarrow \\ \rightarrow Exseq(x, t) \ \& \ (z, y) \in t).$$

Proof. The sequence with a single element $(1, 0)$ has number 30. Further we need the inequality

$$(*) \quad z^{16} \geq \bar{9}(\bar{4}z^4 + 1)^2, \quad \text{for } z \geq \bar{2}.$$

We prove the lemma by induction over y with x and s as parameters. If $y = 0$, then $z = 1$, hence if the sequence t consists of a single element $(1, 0)$, we have $t \leq \bar{30}z^{16}$. Now assume that $y' > 0$ and for every $z \leq s$ there exists a t satisfying the condition. We want to find t' for y' and some z' such that $(z', y') \in s$. By the definition of $Exseq$ there exists $(z, y) \in s$ such that either

$$(i) \quad z' = z^2 \text{ and } y' = \bar{2}y,$$

or

$$(ii) \quad z' = z^2x \text{ and } y' = \bar{2}y + \bar{1}.$$

By the induction assumption we have some t for (z, y) . By Lemma 3.7 we can extend t to t' by adding (z', y') to it and so that

$$t' \leq \bar{9}t((z', y') + \bar{1})^2.$$

Clearly t satisfies $Exseq(x, s)$. By lemma 3.10, we have

$$t' \leq \bar{9}t(\bar{4}(z')^2 + \bar{1})^2.$$

If $z = 1$ then $y = 0$, hence $y' = 1$, $z' = x \geq 2$, and we have, by $(*)$,

$$t' \leq \bar{9} * \bar{30} * (\bar{4}(z')^4 + \bar{1})^2 \leq \bar{30}(z')^{16}.$$

If $z \geq 2$, then in case (i) we have, by $(*)$,

$$t' \leq \bar{9} * \bar{30}z^{16} * (\bar{4}z^4 + \bar{1})^2 \leq \bar{30}z^{32} = \bar{30}(z')^{16},$$

and in case (ii) we have

$$t' \leq \bar{9} * \bar{30}z^{16} * (\bar{4}z^4x^2 + \bar{1})^2 \leq \bar{30}z^{32}x^{16} = \bar{30}(z')^{16}. \quad \square$$

Now we are ready to define the exponentiation.

3.12 Definition.

$$\begin{aligned} \text{Exp}(x, y, z) \equiv & (z = \bar{1} \& y = \bar{0}) \vee (x \leq \bar{1} \& z = x \& y > \bar{0}) \\ & \vee (\exists s \leq \bar{30}z^{16})(\text{Exseq}(x, s) \& (z, y) \in s). \end{aligned}$$

We have written the formula so that it is bounded, however Lemma 3.11 permits us to drop the bound at the existential quantifier, which will be useful below. It is easy to check that the formula satisfies (c.1). To prove (c.2) we first derive the recurrent formulae which correspond to the recursion on notation and then derive (c.2) from them.

3.13 Lemma.

- (c.3) $\text{Exp}(x, y, z) \rightarrow \text{Exp}(x, \bar{2}y, z^2) \& \text{Exp}(x, \bar{2}y + \bar{1}, z^2x);$
 (c.4) $\text{Exp}(x, \bar{2}y, z) \rightarrow (\exists w \leq z)(w^2 = z \& \text{Exp}(x, y, w));$
 (c.5) $\text{Exp}(x, \bar{2}y + \bar{1}, z) \rightarrow (\exists w \leq z)(w^2x = z \& \text{Exp}(x, y, w)).$

Proof. (c.3) The lemma is trivial for $x \leq 1$ or $y = 0$. Suppose $x > 1$, $y > 0$ and $\text{Exp}(x, y, z)$. Let s be a witnessing sequence for $\text{Exp}(x, y, z)$ i.e. $\text{Exseq}(x, s)$ and $(z, y) \in s$. By Lemma 3.7, we can extend this sequence to s_1 (s_2 respectively) by adding another element $(z^2, 2y)$ ($(z^2x, 2y + 1)$ respectively). Then we have $\text{Exseq}(x, s_i)$, for $i = 1, 2$. Now we do not have to worry about the size of s_1 and s_2 , since, by Lemma 3.12, we can replace s_1 and s_2 by suitably bounded ones.

(c.4) Again we can assume $x > 1$, $y > 0$. Let s be a witness for $\text{Exp}(x, \bar{2}y, z)$. Then by definition of Exseq we have some w such that $z = w^2$ and $(w, y) \in s$. Again Lemma 3.12 takes care of the size of s . The proof of (c.5) is quite similar. $\square$

3.14 Lemma ($I\Sigma_0$).

$$\text{Exp}(x, y + \bar{1}, z) \equiv (\exists v \leq z)(\text{Exp}(x, y, v) \& z = vx).$$

Observe that by Lemma 3.8(ii) this is equivalent to (c.2).

Proof. Suppose that the formula is not true for some x, y and z . By $I\Sigma_0$, we can assume that it holds for every $y' < y$ and every $z' \leq z$. Consider two cases:

(i) y is even, say $y = 2p$. Then, by (c.4) and (c.3),

$$\begin{aligned} \text{Exp}(x, y + \bar{1}, z) &\equiv (\exists w \leq z)(w^2 x = z \ \& \ \text{Exp}(x, p, w)) \\ &\equiv (\exists w \leq z)(\text{Exp}(x, \bar{2}p, w^2) \ \& \ w^2 x = z). \end{aligned}$$

Thus taking $v = w^2$ we get a contradiction.

(ii) y is odd, say $y = 2p + 1$. Then, similarly,

$$\begin{aligned} \text{Exp}(x, y + \bar{1}, z) &\equiv \text{Exp}(x, \bar{2}(p + \bar{1}), z) \\ &\equiv (\exists w \leq z)(w^2 = z \ \& \ \text{Exp}(x, p + \bar{1}, w)). \end{aligned}$$

Now by the induction assumption this is equivalent to

$$(\exists w, u \leq z)(\text{Exp}(x, p, u) \ \& \ ux = w \ \& \ w^2 = z);$$

again by Lemma 3.7, we get

$$(\exists w, u \leq z)(\text{Exp}(x, \bar{2}p + \bar{1}, u^2 x) \ \& \ ux = w \ \& \ w^2 = z);$$

The contradiction is obtained now by taking $v = u^2 x$. □

Now we can sum up what we have proved.

3.15 Theorem. The clauses (c.1) and (c.2) are provable in $I\Sigma_0$ for the Σ_0 formula of Definition 3.13. □

One can derive from (c.1) and (c.2) other natural properties of exponentiation in $I\Sigma_0$ such as, for instance,

$$\begin{aligned} \text{Exp}(x, y_1, z_1) \ \& \ \text{Exp}(x, y_2, z_2) &\rightarrow \text{Exp}(x, y_1 + y_2, z_1 z_2); \\ \text{Exp}(x, y_1, z) \ \& \ \text{Exp}(z, y_2, w) &\rightarrow \text{Exp}(x, y_1 y_2, w); \\ \text{Pow}(z) &\equiv (\exists y) \text{Exp}(\bar{2}, y, z). \end{aligned}$$

Thus it is natural to ask whether the clauses (c.1) and (c.2) are sufficient for deriving all the other properties or we have to use sometimes also the explicit definition of the exponentiation relation. The answer is given by the following easy theorem.

3.16 Theorem. Let E be a new ternary relation symbol. Let $I\Sigma_0(E)$ be the theory $I\Sigma_0$ with the induction schema extended to bounded formulae containing E and with additional axioms

$$\begin{aligned} \text{(c.1)'} \quad & E(x, \bar{0}, z) \equiv z = \bar{1}; \\ \text{(c.2)'} \quad & E(x, y + \bar{1}, z) \equiv (\exists v)(E(x, y, v) \ \& \ z = vx). \end{aligned}$$

Then $I\Sigma_0(E)$ proves

$$E(x, y, z) \equiv \text{Exp}(x, y, z).$$

Thus whatever we can prove about the exponentiation in $I\Sigma_0$, it can be derived from (c.1) and (c.2) (using the fact that it is defined by a bounded formula).

Proof. By induction over y we prove the following formula in $I\Sigma_0(E)$:

$$(\forall z \leq u)(E(x, y, z) \equiv \text{Exp}(x, y, z)). \quad \square$$

Let us note that what we have described is not the only possible approach. In fact it is even possible to find a Σ_0 formula for exponentiation which does not use any concept of coding sequences, based on the representation of the power x^y in the form

$$x^y = (x - 1)[(x - 1)v + y] + 1,$$

for some v , see [Pudlák 83, A definition].

(d) Developing $I\Sigma_0 + \Omega_1$

For a moment we shall digress to $I\Sigma_0 + \Omega_1$. Recall that the theory $I\Sigma_0 + \Omega_1$ is the theory in the language L_0 with the following axioms

$$Q + I\Sigma_0 + \Omega_1$$

where Ω_1 is

$$(\forall x)(\exists y)(y = \omega_2(x)).$$

Let us note that often another function $\omega(x) = x^{|x|}$ is used instead of ω_2 .

The function ω_2 can be expressed using the exponentiation and the length functions as follows.

$$\omega_2(x) = 2^{2^{(|x|-1)-1}}$$

Thus formula $y = \omega_2(x)$ is just an abbreviation which uses the definition of exponentiation and the length function. Hence in order to be able to state Ω_1 we must first construct a definition of exponentiation. If we want to work only in $I\Sigma_0 + \Omega_1$ and do not need the definition of exponentiation in $I\Sigma_0$, we can use a simpler definition of exponentiation.

3.17 Definition.

$$\begin{aligned} \text{Exp}_1(x, y, z) \equiv & (x = \bar{0} \ \& \ z = \bar{0}) \vee (x = \bar{1} \ \& \ z = \bar{1}) \\ & \vee (\exists s)[\text{Seq}(s) \ \& \ (z, y) \in s \ \& \ (\forall v \leq s)((v, \bar{0}) \in s \equiv v = \bar{1}) \\ & \ \& \ (\forall i < y)(\forall w < s)((w, i + \bar{1}) \in s \equiv (\exists v \leq s)((v, i) \in s \ \& \ w = vx))]. \end{aligned}$$

Here the defining formula is not bounded. One can show that s can be bounded by a polynomial in $\omega_2(z)$, and that there is no better bound. This has the following effect. If, for instance, we define the length function, which we need for the axiom Ω_1 , by

$$|x|_1 = y \equiv (\exists z)(Exp_1(\bar{2}, y, z) \& x + \bar{1} \leq z < \bar{2}(x + \bar{1})),$$

then already the statement

$$(\forall x)(\exists y)(y = |x|_1)$$

implies Ω_1 . (Of course, this would be an awkward way to define $I\Sigma_0 + \Omega_1$.) The bound to the size of s in $Exp_1(x, y, z)$ can be proved in $I\Sigma_0 + \Omega_1$. Since induction for formulae with such a bound is also provable in $I\Sigma_0 + \Omega_1$ (Proposition 1.3), $Exp_1(x, y, z)$ is equivalent to $Exp(x, y, z)$, (and to any bounded definition satisfying the inductive clauses).

The simplified definition of exponentiation in $I\Sigma_0 + \Omega_1$ does not save us much work, since we still have to prove a lot of things. But for further concepts Ω_1 helps very much. Most of the syntactical concepts are naturally defined by recursion. The natural way to formalize definitions by recursion is to state that there exists a sequence (or a set) that codes the preceding values of the function. (This is called the course of values definition.) These sequences are usually of polynomial length, which is just captured by Ω_1 . To get around with sequences of linear length requires additional tricks. We shall mention simpler definitions of the function *number of ones* and the property *being a term* in $I\Sigma_0 + \Omega_1$ in subsections (e) and (g).

(e) The Number of Ones in a Binary Expansion

We need another important Σ_0 definable function, the number of occurrences of 1 in the binary expansion of a number x . We shall denote this function by $nuon(x)$. As $nuon(x) \leq x$, the function is provably total in $I\Sigma_0$. Using complexity considerations, it is easy to show that $nuon$ is Σ_0 definable, since it is computable in *LogSpace*. However, again, we need a Σ_0 definition which captures its properties in $I\Sigma_0$, thus we have to define it explicitly.

The basic properties of $nuon(x)$ are the following:

- (e.1) $nuon(\bar{0}) = \bar{0}$;
- (e.2) $nuon(\bar{2}x) = nuon(x)$;
- (e.3) $nuon(\bar{2}x + \bar{1}) = nuon(x) + \bar{1}$.

Again, one can easily show in $I\Sigma_0$ that any two Σ_0 definitions satisfying the clauses above are equivalent.

Our next goal is to define $nuon(x)$ by a Σ_0 formula and prove (e.1–3) in $I\Sigma_0$. We shall proceed faster than in the previous section. From now on we shall use the usual notation for exponentiation instead of $Exp(x, y, z)$. Thus if x^y is used in a formula without any comment, it means that the existence of this number is somehow guaranteed.

3.18 Definition.

- (i) $sgm(x, i, j) = y \equiv (\exists x_1, x_2 \leq x)(x = x_1 2^j + y 2^i + x_2 \ \& \ x_2 < 2^i \ \& \ y < 2^{j-i})$ for $i < j$,
otherwise 0;
- (ii) $bit(x, i) = sgm(x, i, i + \bar{1})$;
- (iii) $|x| = \min\{y \mid x < 2^y\}$.

Here $sgm(x, i, j)$ is the part of x between the i -th and $(j - 1)$ -th digit; $bit(x, i)$ is the i -th dyadic digit of x (see I.1.31); $|x|$ is the length of the binary representation of x . (The length function has been defined before, formula (iii) gives an explicit definition of it in terms of the exponentiation.) We leave to the reader to check that formula (ii) is equivalent to a Σ_0 formula and that $2^{|x|}$ has the same meaning as $lpw(x)$.

3.19 Definition.

$$\begin{aligned} Nuonseq(s, x, y) \equiv Seq(s) \ \& \ (\forall z \leq s)((z, \bar{0}) \in s \equiv z = \bar{0}) \ \& \ (y, |x|) \in s \ \& \\ \& \ (\forall i < |x|)(\forall z \leq s)((z, i + \bar{1}) \in s \equiv (\exists v \leq s)((v, i) \in s \\ \& \ z = v + bit(x, i))) . \end{aligned}$$

3.20 Lemma. There exists an absolute constant K such that $I\Sigma_0$ proves

$$2^{|x|^2} \text{ exists} \rightarrow (\exists s \leq 2^{K * |x|^2})(\exists y \leq x) Nuonseq(s, x, y) .$$

(The antecedent should be written more precisely as

$$(\exists z) Exp(\bar{2}, |x|^2, z) .$$

Proof. The formal proof in $I\Sigma_0$ would be similar to the proof of Lemma 3.11. Here we only check the size of s in the standard model. The sequence s contains $|x|$ elements of the form (z, i) where $z \leq i \leq |x|$. Thus its length is at most

$$K * |x| * \|x\| \leq K * |x|^2 . \quad \square$$

3.21 Lemma ($I\Sigma_0$).

$$\begin{aligned} Nuonseq(s, x, y) \ \& \ Nuonseq(t_1, \bar{2}x, y_1) \rightarrow y_1 = y ; \\ Nuonseq(s, x, y) \ \& \ Nuonseq(t_2, \bar{2}x + \bar{1}, y_2) \rightarrow y_2 = y + \bar{1} . \end{aligned}$$

Proof. Using induction over $i = 0, \dots, |x|$ show that

$$\begin{aligned}(z, i + \bar{1}) \in t_1 &\equiv (z, i) \in s; \\ (z, i + \bar{1}) \in t_2 &\equiv (z, i + \bar{1}) \in s.\end{aligned}\quad \square$$

In $I\Sigma_0 + \Omega_1$ we can define $nuon(x) = y$ by

$$nuon_1(x) = y \equiv (\exists s) Nuonseq(s, x, y),$$

since by Lemma 3.20 we know that s can be bounded by a function which is total in $I\Sigma_0 + \Omega_1$, (the subscript is used to distinguish different formulae). Then we can use Lemma 3.21 to prove clauses (e.1–e.3). Unfortunately we cannot use this formula in $I\Sigma_0$ as there is no polynomial bound to such an s . We have to use a more complicated construction. The idea is similar to that of Nepomnjaščij presented in Sect. 1 of this chapter. Namely, we cut x into segments x_i such that on each segment x_i we can use $nuon_1(x_i)$, because the witnessing sequences are bounded by a polynomial in x (i.e. the pieces are small enough). Then $nuon(x)$ is computed as the sum of the values $nuon_1(x_i)$.

We need the following function, whose graph is Σ_0 definable in $I\Sigma_0$:

$$\begin{aligned}lf(x) &= 0 && \text{if } x = 0, \\ &= |x^x| && \text{otherwise.}\end{aligned}$$

3.22 Definition. Let $nuon(0) = 0$ and, for $x > 0$, let

$$\begin{aligned}nuon(x) = y &\equiv (\exists t) [Seq(t) \& (\forall z \leq t) [(z, \bar{0}) \in t \equiv z = \bar{0}] \\ &\& (\forall i \leq x) (lf(i) < |x| \leq lf(i + \bar{1}) \rightarrow (y, i + \bar{1}) \in t) \\ &\& (\forall i \leq x) (\forall z \leq t) [lf(i) < |x| \rightarrow ((z, i + \bar{1}) \in t \\ &\equiv (\exists v \leq t) ((v, i) \in t \& \\ &\& z = v + nuon_1(sgm(x, lf(i), lf(i + 1)))))]].\end{aligned}$$

In order to show that this formula is Σ_0 over $I\Sigma_0$ we have to find a polynomial bound for t and for the sequences s in $nuon_1$. First observe that, for a sufficiently large constant c ,

$$i \geq c \frac{|x|}{\|x\|} \rightarrow |i^i| \geq |x|.$$

Thus we need only those i 's which are smaller than this bound. Now we have to estimate

$$sgm(x, lf(i), lf(i + 1)).$$

The length of this number is bounded by $lf(i+1) - lf(i)$. Using the inequality $a < 2^{|a|} \leq 2a$, for $a > 0$, we get, for $i > 0$,

$$2^{|(i+1)^{i+1} - i^i|} = \frac{2^{|(i+1)^{i+1}|}}{2^{i^i}} < \frac{2^{(i+1)^{i+1}}}{i^i} = 2(i+1)(1 + \frac{1}{i})^i,$$

which tends to $2e(i+1)$ as $i \rightarrow \infty$. Thus we have, for some constants c', c'' and $i > 0$,

$$(e.4) \quad \begin{aligned} sgm(x, lf(i), lf(i+1)) &\leq 2^{lf(i+1) - lf(i)} = 2^{|(i+1)^{i+1} - i^i|} \\ &\leq c'(i+1) \leq c'' \frac{|x|}{\|x\|}. \end{aligned}$$

Now, by Lemma 3.20, $nuon_1(sgm(x, lf(i), lf(i+1)))$ is witnessed by a sequence s such that

$$s \leq 2^{K|sgm(x, lf(i), lf(i+1))|^2} \leq 2^{K|c'' \frac{|x|}{\|x\|}|^2} \leq 2^{K'|x|},$$

for some constant K' , which is a bound polynomial in x . The sequence t consists of $c \frac{|x|}{\|x\|}$ elements. The elements have form (z, i) with

$$i \leq c \frac{|x|}{\|x\|}, \quad z \leq |x|,$$

thus the length of (z, i) is linear in $\|x\|$. Hence the length of t is linear in $|x|$, i.e. t is bounded by a polynomial in x .

The proof above has been done in the standard model. The proof in $I\Sigma_0$ would be more complicated, but very similar to the proof of Lemma 3.11. We sketch briefly this argument.

First we can estimate in $I\Sigma_0$

$$(1 + 1/i)^k \leq 1 + 2k/i + 2k^2/i^2,$$

for $k \leq i$, hence we get $(1 + 1/i)^i \leq 5$ and we can use (e.4). Secondly we need to estimate elements (z, i) of the sequence t for $|i^i| \leq |x|$. This is done by showing that $nuon_1(a) \leq |a|$, whenever defined (by induction on the length of a), from which we get $(z, i) \leq p(|x|)$ for some polynomial p . For a given x take the largest j such that $|j^j| \leq |x|$. By induction over j we show that the sequence t can be bounded by $q(i^i)$ for a suitable polynomial which we shall specify later. Suppose that such a bound holds for j and let x be such that $j+1$ is the maximal such that $|(j+1)^{j+1}| \leq |x|$. Let $x' = sgm(x, \bar{0}, j^j)$. Thus we have some t' for x' with $t' \leq q(j^j)$. To obtain t for x , we only need to add

an element $(z, j+1)$ to t . Since $(z, j+1) \leq p(|x'|)$ and $j+1$ is about $\frac{|x'|}{|x''|}$, we have, for a suitable polynomial r , $(z, j+1) \leq r(j)$. Thus, by Lemma 3.7

$$t \leq t' * 9(r(j) + 1)^2 \leq q(j^j) * r'(j).$$

The induction can now be completed, if we ensure that

$$q(j^j) * r'(j) \leq q((j+1)^{j+1}),$$

which can be done by taking polynomial q sufficiently large with respect to r' .

Thus we have shown:

3.23 Lemma. In $I\Sigma_0$ the formula defining $nuon(x) = y$ is equivalent to a Σ_0 formula. $\square$

Remark. There are, of course, other possible ways to divide the binary expansion of x for the definition of $nuon$. The reasons why we have chosen the partition determined by the function $lf(i)$ are the following:

- (1) it does not depend on x , hence we can easily use induction;
- (2) it grows smoothly, (while e.g. $i * |i|$ makes big jumps).

3.24 Lemma.

- (i) The formula $nuon(x) = y$ defines a function in $I\Sigma_0$.
- (ii) The formula $nuon(x) = y$ satisfies (e.1–e.3) provably in $I\Sigma_0$.

Proof. The proof of (i) is omitted. The proof of (e.1) follows from the definition. To prove (e.2) and (e.3) we would like to simulate the proof of Lemma 3.21. Now we cannot code the sequence of the values of $nuon(x')$ for all initial segments x' of x by a number, but we can define these values by a Σ_0 formula. Namely, let t be a witness of $nuon(x) = y$, let $k \leq |x| + 1$. Then the value $z = nuon(x')$ for the initial segment x' of x of length k is defined by

$$\begin{aligned} \varphi(z, k, t, x) \equiv & (\exists i, v)(|i^i| < k \leq |(i+1)^{i+1}| \& (v, i) \in t \\ & \& z = v + nuon_1(sgm(x, |i^i|, k))). \end{aligned}$$

Similarly as above one can show that φ is Σ_0 in $I\Sigma_0$. Now let t' be a witness for $nuon(2x) = y'$. Since, by Lemma 3.21, $nuon_1(2a) = nuon_1(a)$, we have, by induction on k ,

$$\varphi(z, k, t, x) \equiv \varphi(z, k, t', 2x).$$

But we have also

$$\begin{aligned} \varphi(z, |x| + 1, t, x) & \equiv z = y, \\ \varphi(z, |2x|, t', 2x) & \equiv z = y', \end{aligned}$$

hence $y = y'$. The proof of (e.3) is similar. $\square$

3.25 Lemma.

$$nuon(x2^{|y|} + y) = nuon(x) + nuon(y).$$

Proof. First we show that

$$\begin{aligned} x2^{|2y|} + 2y &= 2(x2^{|y|} + y), \\ x2^{|2y|} + 2y + 1 &= 2(x2^{|y|} + y) + 1, \end{aligned}$$

and then use (e.1–e.3) and induction *PIND* (which is derivable from the ordinary induction, see Sect. 4) over y . $\square$

Also here there is an alternative way to define the concept *nuon*. This was suggested by S. Buss, (unpublished). The idea is to count ones in the straightforward way as in *nuon*₁ except that we include in the sequence of *Nuonseq* only the bits which have been changed. Thus e.g. the following sequence

$$1, 10, 10, 11, 100, 101, 101, 110, 111, 1000,$$

will be replaced by

$$1, 10, \quad , 1, 100, 1, \quad , 10, 1, 1000.$$

Then we need at most $2n$ bits to count up to the number n (plus bits needed to separate the members of the sequence).

(f) Coding Sequences

Using the function *nuon* we are now able to define a very efficient coding, which is close to the information theoretical lower bound. Before defining *nuon* we were able to say that an element x belongs to a sequence and that an element x precedes an element y in sequence s . Now we can define that x is the i -th element of s .

3.26 Definition.

$$\begin{aligned} (s)_i &= x \text{ if } (\exists u, v \leq s)(\exists y, y' \leq v)(\exists u_1, v_1 < y)(\exists u_2, v_2 \leq v)(Seq(s) \\ &\quad \& s = (u, v) \& Pow(y) \& Pow(y') \& y' \geq \bar{2} \& x < y' \\ &\quad \& u = u_2 y' y + x y + u_1 \& v = v_2 \bar{2} y' y + y' y + y + v_1 \\ &\quad \& i = nuon(v_1)). \\ &= \bar{0} \text{ otherwise.} \end{aligned}$$

Thus $(s)_i = x$ is defined by the same formula as $x \in s$ (Definition 3.6) except that we have added the condition $i = nuon(v_1)$. The meaning of this condition

is clear: i is the number of markers before the beginning of the occurrence of x in u . Next we define the length of a sequence and concatenation of two sequences.

3.27 Definition.

- (i) $lh(s) = nuon(v) - 1$, if $Seq(s) \ \& \ (\exists u \leq s) s = (u, v)$
 $= 0$ otherwise.
 (ii) For s, t such that $Seq(s), Seq(t), s = (u, v)$ and $t = (u', v')$

$$s \frown t = (u'', v''),$$

where

$$u'' = u'2^{|v|-1} + u,$$

$$v'' = (v' - 1)2^{|v|-1} + v;$$

otherwise we set $s \frown t = \bar{0}$.

Note that these functions are defined by Σ_0 formulae. We shall state the most important properties of concatenation. The role of the empty sequence is played by $(1, 0)$.

3.28 Lemma ($I\Sigma_0$). For all x, y, z such that $Seq(x), Seq(y)$ and $Seq(z)$

- (i) $(\bar{1}, \bar{0}) \frown x = x \frown (\bar{1}, \bar{0}) = x$;
 (ii) $x \frown y = x \frown z \rightarrow y = z$;
 (iii) $y \frown x = z \frown x \rightarrow y = z$;
 (iv) $x \frown (y \frown z) = (x \frown y) \frown z$;
 (v) $lh(\bar{1}, \bar{0}) = \bar{0}$, $lh(x \frown y) = lh(x) + lh(y)$.

Proof. The proofs are just tedious verifications. Therefore we prove only (ii) to illustrate the proof technique.

Let $x = (u(x), v(x))$, $y = (u(y), v(y))$, $z = (u(z), v(z))$. Suppose $x \frown y = x \frown z$, then

$$(v(x) - 1)2^{|v(y)|-1} + v(y) = (v(x) - 1)2^{|v(z)|-1} + v(z).$$

Since the function $a \mapsto 2^{|a|-1}$ is nondecreasing, we have $v(y) = v(z)$. Further we have

$$u(x)2^{|v(y)|-1} + u(y) = u(x)2^{|v(z)|-1} + u(z).$$

Since, by the definition of Seq , $|u(y)| < |v(y)|$, $|u(z)| < |v(z)|$ and since $v(y) = v(z)$, we have $u(y) = u(z)$ by the lemma about division with remainder Theorem I.1.10 (6). $\square$

Since $x, y \mapsto x \frown y$ is a total function in $I\Sigma_0$, it must be bounded by a polynomial in x, y . We shall state an explicit bound now. The proof is just an easy computation.

3.29 Proposition ($I\Sigma_0$). $s \frown t \leq \overline{64}st$.

The maximal element of a sequence s is trivially definable by a Σ_0 formula, it will be denoted by $\max(s)$. We shall use $\max(s)$ and $lh(s)$ to write down an explicit bound to the size of t which codes the same sequence of numbers as s . We cannot bound s itself, since numbers in s can be represented by 0-1 sequences with many 0's before the first 1. This is a drawback of our definition.

3.30 Proposition.

$$\begin{aligned} Seq(s) \rightarrow (\exists t)(Seq(t) \& lh(t) = lh(s) \& (\forall 0 \leq i < lh(s))((t)_i = (s)_i) \\ \& |t| \leq (2|\max(s)| + 4) * lh(s) + 2). \end{aligned}$$

Proof. By induction on j , $j = lh(s)$, $lh(s) - 1, \dots, 0$, we shall prove the following formula

$$\begin{aligned} (\exists t)(Seq(t) \& lh(t) = lh(s) - j \& (\forall 0 \leq i < lh(s) - j)((t)_i = (s)_{j+i}) \\ \& |t| \leq (2|\max(s)| + 4) * (lh(s) - j) + 2). \end{aligned}$$

For $j = lh(s)$ put $t = (0, 1) = 1$. Suppose it is true for $0 < j \leq lh(s)$, i.e. we have some t' for j . Then let t be t' extended by $(s)_{j-1}$ as described in the proof of Lemma 3.7. It is not difficult to check that $(t)_i = (t')_{i-1}$ for $i = 1, \dots, lh(s) - j$, and $(t)_0 = (s)_{j-1}$. Hence $(t)_i = (s)_{j-1+i}$. Further we have

$$t \leq 9t'((s)_{j-1} + 1)^2 \leq 9t'(\max(s) + 1)^2.$$

An easy computation gives us

$$|t| \leq 4 + |t'| + 2|\max(s)| \leq (2|\max(s)| + 4)(lh(s) - j + 1) + 2.$$

Thus the formula is proved. Taking $j = 0$ we obtain our proposition. $\square$

Suppose a finite sequence of numbers is defined by a Σ_0 formula. In $I\Sigma_0 + Exp$ such a sequence always has a code. When does it have a code in $I\Sigma_0$, i.e. which comprehension principles hold in $I\Sigma_0$? The answer is that there is only one essential restriction: the length of the code must be linear in the parameters. We state an example of a replacement schema which is valid in $I\Sigma_0$.

3.31 Proposition. For every Σ_0 formula $\varphi(x, y)$ $I\Sigma_0$ proves

$$\begin{aligned} (\forall x \leq u)(\exists y \leq v)\varphi(x, y) \& (\exists z)(z = (v + 2)^u) \rightarrow \\ \rightarrow (\exists s)(lh(s) = u \& (\forall x < u)(\varphi(x, (s)_x) \& (s)_x \leq v)). \end{aligned}$$

Proof. By induction on u ; similar to the proof of Proposition 3.30. $\square$

(g) Syntactical Concepts

We shall use natural representation of syntactical objects as particular sequences (words) in a finite alphabet. Since we have concatenation, we do not have problems with the properties which can be derived from definitions based on rudimentary formulae. Recall that x is a *part* of y is defined by

$$x \subseteq_p y \equiv (\exists u, v \leq y)(y = u \frown x \frown v),$$

which is a Σ_0 formula. (Rudimentary formulae are the formulae in the language $\{=, \frown\}$ with bounded quantifiers of the form $\exists x \subseteq_p y$ and $\forall x \subseteq_p y$.) Thus, for instance, once we have the concept of a formula, we can define the concept of being a subformula and prove its properties. Similarly there are no problems with defining that a variable does (not) occur in a formula etc.

What is not quite obvious is that we can define the concept of terms and formulae by Σ_0 formulae. Actually, it is sufficient to define only terms, since terms are essentially trees and other syntactical concepts can be represented as particular labelled trees with some additional properties (these properties usually defined by rudimentary formulae).

We shall consider a simplified situation: We have four symbols, a variable v , brackets $(,)$, and a binary operation symbol $\circ$. Then the terms are the smallest set satisfying:

- (1) v is a term;
- (2) if s and t are terms, then $(s \circ t)$ is a term.

A formula $Term(x)$ defining the concept of terms must satisfy the following clauses:

- (g.1) $Term(v)$;
- (g.2) $Term(x) \& Term(y) \rightarrow Term((\frown x \frown \circ \frown y \frown))$;
- (g.3) $Term(x) \rightarrow x = v \vee (\exists p, q)(Term(p) \& Term(q) \& x = (\frown p \frown \circ \frown q \frown))$.

The first two express formally (1) and (2), (g.3) corresponds to the clause that " $Term$ is the smallest set satisfying (1) and (2)". In fact we can prove that any two Σ_0 formulae satisfying (g.1–3) are equivalent in $I\Sigma_0$. It is exactly the same situation as with exponentiation and function $nuon$ of subsection (e).

In $I\Sigma_0 + \Omega_1$ we can use the following simple definition:

$$Term_1(x) \equiv (\exists s)(Seq(s) \& x \in s \& (\forall y \in s)(y = v \vee (\exists p, q)(p \in s \& q \in s \& y = (\frown p \frown \circ \frown q \frown))))).$$

From now on we agree to omit the sign for concatenation whenever there is no danger of confusion. The proof of (g.1–3) for $Term_1$ in $I\Sigma_0 + \Omega_1$ is easy.

Our approach in $I\Sigma_0$ will be based on counting brackets. The occurrences of brackets (or other symbols) can be easily coded by occurrences of ones in the binary expansion of a number. Thus we can again use function *nuon*. Since this transformation is easy, we shall use less formal descriptions of formulae.

3.32 Definition. (well-bracketing, sub-well-bracketing, term)

- (i) $WB(x) \equiv$ (1) the number of "(" in x equals to the number of ")" in x ;
 (2) the number of ")" in each proper initial segment of x is less than the number of "(" in it;
 (3) x is nonempty.
- (ii) $SubWB(x, y) \equiv WB(x) \& WB(y) \& x \subseteq_p y$.
- (iii) $Term(x) \equiv WB(x) \& (\forall y)(SubWB(y, x) \rightarrow y = \mathbf{v} \vee (\exists p, q)(SubWB(p, x) \& SubWB(q, x) \& y = (p \circ q)))$.

3.33 Lemma. $WB(x)$, $SubWB(x, y)$ and $Term(x)$ are Σ_0 over $I\Sigma_0$.

Proof. These formulas use Σ_0 defined predicates and functions (for counting brackets use the function *nuon*); the quantification can be made bounded. $\square$

3.34 Lemma. $I\Sigma_0$ proves (g.1–3) for the definition of $Term(x)$ above.

Proof. (g.1) is obvious. To prove (g.2), assume $Term(x)$ and $Term(y)$ and let $z = (x \circ y)$. The condition $WB(z)$ is easy. Let u be a sub-well-bracketing of z . Clearly it is now sufficient to show that u is either a part of x or a part of y or u equals to z . If u is not a part of x nor a part of y , it must contain the first occurrence of "(" or the occurrence of $\circ$ between x and y or the last occurrence of ")".

(1) Suppose u contains the first occurrence of "(", then it must contain also the last occurrence of ")", otherwise condition (1) of $WB(u)$ would be violated.

(2) The case with the last occurrence of ")" is symmetric.

(3) Suppose u contains $\circ$ between x and y . Then u is not $\mathbf{v}$, hence it contains an end segment of x or an initial segment of y . In the first case it must contain the first occurrence of "(" and in the second case it must contain the last occurrence of ")", since otherwise condition (2) of $WB(u)$ would be violated. Hence again $u = z$ and condition (g.2) is proved.

Now we prove (g.3). Suppose $Term(x)$. If $x = \mathbf{v}$, then there is nothing to prove. If not, then $x = (p \circ q)$ for some p, q such that $SubWB(p, x)$ and $SubWB(q, x)$. To prove that p and q are terms, we have to check that each sub-well-bracketing y of p or of q has the required form. But each such y is also a sub-well-bracketing of x , hence by definition it must have this form. $\square$

Note that we can think of *WB* as a definition of arbitrary finite trees and *Term* as a definition of finite binary trees. The last property of terms that we prove is sometimes called *unique readability*.

3.35 Lemma ($I\Sigma_0$).

$$\text{Term}(x) \& \text{Term}(y) \& \text{Term}(p) \& \text{Term}(q) \& (x \circ y) = (p \circ q) \rightarrow \\ \rightarrow x = p \& y = q.$$

Proof. Assume the antecedent is true. Then p and q are sub-well-bracketings of $(x \circ y)$. It follows now from the proof of Lemma 3.34 that neither p nor q can contain the occurrence of $\circ$ between x and y . Hence the only possibility is that $x = p$ and $y = q$. $\square$

Now we consider *substitution*. We have defined terms in a very restricted language; the following proposition holds for terms in the usual general sense.

3.36 Proposition. The relations *the term u is the result of substituting s for all (some, respectively) occurrences of the variable v in t* are Σ_0 definable.

Proof. We need to express that t is of the form

$$p_1 \frown v \frown p_2 \frown v \frown \dots \frown p_{n-1} \frown v \frown p_n,$$

where v does not occur (may occur, respectively) in $p_1, p_2, \dots, p_n$, and u is of the form

$$p_1 \frown s \frown p_2 \frown s \frown \dots \frown p_n \frown s \frown p_n.$$

This can be done by introducing two 0-1 sequences w_1, w_2 such that w_1 marks occurrences of v in t and w_2 marks occurrences of s in u (not necessarily all). Then we state that the word between the i -th and $i+1$ -st marked occurrence of v in t is equal to the word between the i -th and $i+1$ -st marked occurrence of s in u . To express that an occurrence is the i -th one we use, of course, our function *nuon*. $\square$

Since the definition of substitution is quite straightforward, the usual properties are easily provable in $I\Sigma_0$ and we shall not present them here. In $I\Sigma_0 + \Omega_1$ we can define the total ternary operation

$$t, v, s \mapsto t(v/s),$$

where $t(v/s)$ stands for the result of substituting s for all occurrences of variable v . This is because we can bound the length of $t(v/s)$:

$$|t(v/s)| \leq \text{const} * |t| * |s|.$$

Since the bound is, essentially, the best possible and it is not linear, this function is not total in $I\Sigma_0$. This causes some problems, when we develop metamathematics in $I\Sigma_0$. For example, we cannot replace occurrences of a variable by a term in a proof.

The purpose of this section is to study combinatorial concepts needed for formalization. It should be clear by now that these concepts are quite simple and can be handled in $I\Sigma_0$, unless they are functions growing more than polynomially. In the next section we prove a general theorem which covers some standard cases.

Let us consider the concept of a proof as an example. Suppose we formalize proofs as sequences of formulae. Then we need to write down that each element of the sequence is a formula which follows from the preceding ones by a rule. The relation that a rule defines on formulae, is usually constructed from notions whose definability and properties have been shown in $I\Sigma_0$ (concatenation, substitution, occurrence etc.). On the other hand the provability predicate which we obtain in this way does not have the basic closure properties in $I\Sigma_0$. This is because it is a Σ_1 concept.

(h) Formalizations Based on Context-Free Grammars

After working for some time in a formal system one gets a feeling, usually quite a clear one, of what can be formalized. Repeating the same tricks in formalizations again and again becomes boring. Then comes a dream: to prove a general theorem which would take care of all situations. Usually it remains only a dream, because if we want to state a comprehensible and memorable theorem, we have to give up a lot of its possible applications. We are going to state and prove a theorem which is such a compromise.

We shall use the concept of a context-free grammar. A *grammar* is determined by two alphabets and rewriting rules. A *rewriting rule* is a pair of words (u, v) . By an application of the rule (u, v) to a word w we mean the replacement of an occurrence of u in w by v . The rules can be applied in arbitrary order to arbitrary occurrences of subwords. One of the alphabets consists of so called *nonterminal symbols* the other consists of *terminal symbols*. We are interested in the words which consist of terminal symbols only and which can be derived from a specified initial nonterminal symbol using rewriting rules. This is *the language* determined by the grammar. A *context-free grammar* is a grammar in which the first word of the rewriting rule is always a single symbol. In context-free grammars we can interpret nonterminal symbols as *concepts* (e.g. "noun", "verb" in natural languages, or "term", "procedure" in programming languages). The rules can be thought of as a kind of inductive clauses, which determine the structure of the words in the language. We know that it is important to have proofs of inductive clauses

in $I\Sigma_0$. This is exactly what our theorem will do. Examples will be shown at the end of the section.

Let us state a formal definition of context-free grammars and introduce the standard notation. A *context-free grammar* is a quadruple $G = (N, T, P, S_0)$, where N and T are finite sets, $N \cap T = \emptyset$, $S_0 \in N$, $P \subseteq N \times (N \cup T)^+$, where $+$ denotes nonempty words in a given alphabet. (The standard definition also allows the derivation of empty words, but it would cause us some technical complications, therefore we consider only languages without empty words; the results that we shall prove do not depend on this restriction.) The elements of N (T respectively) are called *nonterminal symbols* (*terminal symbols*, respectively), S_0 is the *initial nonterminal symbol*, P is a set of *rewriting rules*. Let $u, w \in (N \cup T)^+$; we shall write

$$u \Rightarrow w,$$

(or $u \xRightarrow{G} w$ to indicate which grammar we are using), if there are $A \in N$ and $v \in (N \cup T)^+$ such that A occurs in u , $(A, v) \in P$ and w results from u by replacing an occurrence of A by v . The relation $\xRightarrow{*}$ denotes the transitive closure of $\Rightarrow$. Thus $u \xRightarrow{*} w$ means that $u = w$ or u can be rewritten to w by applications of the rules of the grammar. Note that the rewriting process is nondeterministic: we can choose rules and occurrences of letters. The set of all words $w \in T^+$ such that $S_0 \xRightarrow{*} w$ is the *language generated by G* . The languages generated by context-free grammars are called *context-free languages*.

We shall add another unessential *restriction* that there is no rule of the form $(A, B) \in P$ with $A, B \in N$. In order to simplify notation, we shall omit the sign for concatenation in the rest of this section.

Our aim is to formalize the property of being derivable from A ; this is the meaning of $\varphi_A(w)$ below.

3.37 Theorem. Let G be a context-free grammar. Then one can construct a family of Σ_0 formulae φ_A for $A \in N$ with the following property: If

$$A \Rightarrow w_{i1}B_{i1}w_{i2}B_{i2}\dots B_{in_i-1}w_{in_i}, \quad i = 1, \dots, k,$$

are all the rules of G with A on the left side and $w_{i1}, w_{i2}, \dots, w_{in_i} \in T^*$, $B_{i1}, B_{i2}, \dots, B_{in_i-1} \in N$, then $I\Sigma_0$ proves

$$(g.1) \quad \varphi_A(x) \equiv \bigvee_{i=1}^k (\exists y_1) \dots (\exists y_{n_i-1}) (\varphi_{B_{i1}}(y) \& \dots \& \varphi_{B_{in_i-1}}(y) \\ \& x = w_{i1}y_1w_{i2}y_2 \dots y_{n_i-1}w_{in_i}),$$

(where we admit $n_i = 1$).

Proof. Let a context-free grammar $G = (N, T, P, S_0)$ be given. We consider an auxiliary grammar $\underline{G} = (N, \underline{T}, \underline{P}, S_0)$ defined by adding a new terminal $\underline{A}$ for each nonterminal $A \in N$ of G and two new terminal symbols $[,]$, and by replacing each rule of G

$$A \Rightarrow w$$

by

$$A \Rightarrow \underline{A}[w],$$

($\underline{A}[w]$ is the concatenation of $\underline{A}$, $[$, w and $]$.) The idea behind this is that $\underline{G}$ codes the syntactical trees of derivations in G .

The words $w \in (N \cup \underline{T})^+$ which are derivable from some A in $\underline{G}$ can be defined by the following clauses:

- (1) w is a well-bracketing with respect to $[$ and $]$;
- (2) w is of the form $\underline{A}[u]$ for some $A \in N$;
- (3) each occurrence of $[$ is preceded by some $\underline{A}$, for $A \in N$;
- (4) if

$$(g.2) \quad \underline{A}[w_1 \underline{B}_1[v_1]w_2 \dots w_{n-1} \underline{B}_{n-1}[v_{n-1}]w_n]$$

is a sub-well-bracketing of w , then

$$(g.3) \quad A \Rightarrow_{\underline{G}} w_1 \underline{B}_1 w_2 \dots w_{n-1} \underline{B}_{n-1} w_n,$$

with $w_1, \dots, w_n \in T^*$, $\underline{B}_1, \dots, \underline{B}_{n-1} \in N$.

We know that such clauses are expressible by a Σ_0 formula, so we have a Σ_0 formula $\varphi(x)$ which defines words derivable in $\underline{G}$. Moreover the following is provable in IE_0 :

- (5) $\varphi(x) \rightarrow x$ is of the form (g.2) for some rule (g.3);
- (6) $\varphi(\underline{B}_1[v_1]) \& \dots \varphi(\underline{B}_{n-1}[v_{n-1}]) \rightarrow \varphi(\underline{A}[w_1 \underline{B}_1[v_1]w_2 \dots w_{n-1} \underline{B}_{n-1}[v_{n-1}]w_n])$,
for each rule (g.3) of G .

Condition (5) follows immediately from (1), (2) and (4); (6) can be proved in the same way as Lemma 3.34.

Next we need the function which maps words of $(N \cup \underline{T})^+$ on words of $(N \cup T)^*$ so that it only omits the letters which do not belong to $N \cup T$ ($*$ denotes the set of all words, including the empty word). A Σ_0 definition of this function can be constructed using the same idea as we used for substitution. (In fact this function is the substitution of empty words for each $\underline{A}$, $A \in N$ and for each bracket $[$ and $]$.) Let us denote this function by *Forget*(x). It is provable in IE_0 that *Forget* is a homomorphism of the free semigroup $(N \cup \underline{T})^+$ into $(N \cup T)^*$. In particular, it is provable in IE_0 that:

$$(7) \quad \text{Forget}(\underline{A}[w_1 \underline{B}_1[v_1]w_2 \dots w_{n-1} \underline{B}_{n-1}[v_{n-1}]w_n]) \\ = w_1 \text{Forget}(\underline{B}_1[v_1])w_2 \dots w_{n-1} \text{Forget}(\underline{B}_{n-1}[v_{n-1}])w_n.$$

Recall that we do not allow rules of the form (A, B) , for $A, B \in N$. Thus in (g.3) either $n \geq 3$ or at least one w_i is nonempty. This together with (7) can be used to show in $I\Sigma_0$ that:

$$(8) \quad \varphi(x) \rightarrow lh(x) \leq \bar{7} * lh(\text{Forget}(x)),$$

which implies a polynomial bound to x in terms of $\text{Forget}(x)$.

Now we can define our formulae φ_A :

$$(9) \quad \varphi_A(x) \equiv (\exists y)(x = \text{Forget}(\underline{A}y) \& \varphi(\underline{A}y)).$$

Hence by (8) these formulae are Σ_0 over $I\Sigma_0$. The rest of the proof is easy. To prove (g.1), first suppose $\varphi_A(x)$ in $I\Sigma_0$. Then, by (9), for some y and A , $x = \text{Forget}(\underline{A}y)$ and $\varphi(\underline{A}y)$. By (5), $\underline{A}y$ is of the form (g.2). Hence by (7)

$$x = w_1 y_1 w_2 \dots w_{n-1} y_{n-1} w_n,$$

Where

$$y_i = \text{Forget}(\underline{B}_i[v_i]), \quad i = 1, \dots, n-1.$$

Thus, by definition (9), $\varphi_{B_i}(y_i)$ for $i = 1, \dots, n-1$. To prove the converse implication, suppose that

$$x = w_1 y_1 w_2 \dots w_{n-1} y_{n-1} w_n,$$

and

$$\varphi(\underline{B}_1[y_1]) \& \dots \varphi(\underline{B}_{n-1}[y_{n-1}]),$$

for some rule (g.3). Applying (9) and (7) we get $x = \text{Forget}(\underline{A}y)$, for $\underline{A}y$ of the form (g.2). Then, by (6), we have also $\varphi(\underline{A}y)$, thus $\varphi_A(x)$ holds true. $\square$

3.38 Corollary. $A \stackrel{*}{\Rightarrow} w$ iff $\varphi_A(w)$ is true. Hence context-free languages are in Σ_0^N .

Proof. By induction on the length of w . (Repeated application of (g.1) produces a syntactical analysis of w .) $\square$

This corollary is only a by-product. The meaning of the theorem is that it not only produces a Σ_0 formula, but also it gives inductive clauses which determine the concept in $I\Sigma_0$ in a similar way as (c.1-2) determine exponentiation and (e.1-3) determine function $nuon$.

As our first example, we consider the definition of numerals. We cannot use

$$S^n(\bar{0}) = \underbrace{SSS \dots S}_{n \text{ times}}(\bar{0}),$$

in $I\Sigma_0$, since the code of such an expression is exponential in n , thus $I\Sigma_0$ does not prove that $S^n(\bar{0})$ exists for every n . In $I\Sigma_0$ we use *dyadic numerals*

which are defined by

$$\begin{aligned}\bar{0} &= \bar{0}, \quad \bar{1} = S(\bar{0}), \\ \overline{2n} &= (\bar{1} + \bar{1}) * (\bar{n}), \quad \overline{2n+1} = (\bar{1} + \bar{1}) * (\bar{n}) + \bar{1}, \quad \text{for } n > 0.\end{aligned}$$

Using a context-free grammar G we first formalize positive numerals. G has one nonterminal S_0 , terminals $\bar{1}, +, *, (,)$ and rules

$$S_0 \Rightarrow \bar{1} \mid (\bar{1} + \bar{1}) * (S_0) \mid (\bar{1} + \bar{1}) * (S_0) + \bar{1},$$

(this is the standard way of writing three rules at once). Let $\varphi_{S_0}(x)$ be the formula given by Theorem 3.37. Then $I\Sigma_0$ proves

$$\begin{aligned}\varphi_{S_0}(x) \equiv x = \bar{1} \vee (\exists y)(\varphi_{S_0}(y) \& x = '(\bar{1} + \bar{1}) * (y)') \\ \vee (\exists y)(\varphi_{S_0}(y) \& x = '(\bar{1} + \bar{1}) * (y) + \bar{1}'),\end{aligned}$$

where we use '...' to denote a string of characters. Hence, if we define

$$Num(x) \equiv x = \bar{0} \vee \varphi_{S_0}(x),$$

then $Num(x)$ satisfies natural clauses which determine this concept. The value $val(x)$ of a numeral x is easily Σ_0 definable, since the binary expansion of $val(\bar{n})$ is encoded in $\bar{n}$. Thus the function $num(x)$, the x -th numeral, is also Σ_0 definable, namely by

$$num(x) = y \equiv Num(y) \& val(y) = x.$$

In logic we usually reduce the definition of such a concept to terms and thus we do not have to use the full generality of Theorem 3.37. Therefore our second example will be the so called *Polish notation*, which is notation without brackets, thus we cannot reduce it directly to the concept of terms. Let us define variable-free formulae in the language consisting of $\bar{0}, \bar{1}$ – constants, **add**, **mult** – binary operations, **=** – a binary predicate, **non** – a unary connective, **et**, **vel** – binary connectives. The grammar will have two nonterminals S_0, S_1 , the rules are

$$\begin{aligned}S_0 &\Rightarrow =S_1S_1 \mid \text{non}S_0 \mid \text{et}S_0S_0 \mid \text{vel}S_0S_0, \\ S_1 &\Rightarrow \bar{0} \mid \bar{1} \mid \text{add}S_1S_1 \mid \text{mult}S_1S_1.\end{aligned}$$

The meaning of S_0 is “formula”, the meaning of S_1 is “term”. Theorem 3.37 gives us a Σ_0 formula defining variable-free formulae such that the inductive clauses (given by the rewriting rules) are provable in $I\Sigma_0$.

*

We have developed several tools for defining syntactical concepts in $I\Sigma_0$. Their application to the definition of terms and formulae in the full language of arithmetic, to the definition of proofs etc. is quite routine now, so we can end this section here.

4. Witnessing Functions

(a) Introduction

Suppose a sound theory T proves a sentence $(\forall x)(\exists y)\varphi(x, y)$. If T is weak and φ is simple, then we have some additional information on the functions which *witness* the existential quantifier in this sentence, i.e. functions f for which we have $(\forall x)\varphi(x, f(x))$. A classical result says that if T is $I\Sigma_1$ and φ is bounded, then there is such an f which is *primitive recursive*, see Theorem IV.3.6. Parikh [Parikh 71] has noted that f is linear time computable if T is $I\Sigma_0$ and φ is bounded. Buss [Buss 86, Bounded Arith.] has defined a theory S_2 (a conservative extension of $I\Sigma_0 + \Omega_1$), fragments S_2^i and subsets Σ_i^b of bounded formulae in the extended language. His theorems characterize the classes of functions in the polynomial hierarchy using the above phenomenon. In particular, if φ is a Σ_1^b formula and T is S_2^1 , then f is polynomial time computable. The converse of this special case is also true: if f is polynomial-time computable, then there exists a Σ_1^b formula which defines f and for which it is provable in S_2^1 that f is total.

We start this section with the definitions of these classes of formulae and fragments of bounded arithmetic. The next step will be positive results about definability of functions of certain complexity in these fragments. It turns out that this is quite important information about the fragments. For instance, knowing that T_2^i defines functions of $\square_{i+1}^p$, for $i \geq 1$, it is easy to prove Buss' theorem for T_2^i . The main part is subsection (d) where witnessing theorems of Buss and of Krajíček and Takeuti are proved. In subsection (e) we shall show a reduction of the problem of finite axiomatizability of bounded arithmetic to the problem of collapsing the polynomial time hierarchy.

Since the fragments of bounded arithmetic have been extensively studied in [Buss 86, Bounded Arith.], we concentrate on results and proofs not included in that book. In particular, as Buss uses proof theory, we shall give proofs of the witnessing theorems based on model theory. We shall refer to Buss's book for basic facts provable in these fragments; however the readers with some training in bounded arithmetic, for example those who have worked through Sect. 3, should be able to make their own proofs without consulting that book.

(b) Fragments of Bounded Arithmetic

We shall extend the usual language of arithmetic L_0 by adding two unary function symbols $|x|$ and $\lfloor x/2 \rfloor$, and one binary operation $x \# y$. The intended interpretation of $|x|$ is

$|x| = \lceil \log_2(x+1) \rceil$ = the length of the binary expansion of x if $x > 0$, otherwise 0,

i.e. the same as in Sect. 3. Here $\lceil y \rceil$ denotes the least integer $z \geq y$; $\lfloor y \rfloor$ denotes the integer part of y , thus the interpretation of $\lfloor x/2 \rfloor$ is clear. The intended interpretation of the *smash function* is

$x \# y = 2^{|x| * |y|}$ = the power of 2 whose binary expansion has length $|x| * |y| + 1$.

Observe that the set of the lengths of numbers in a sound arithmetical theory containing the smash function is closed under multiplication, hence under any polynomial increase (for standard polynomials). This is important, since it enables us to formalize many standard constructions, in particular polynomial time computations. We shall denote this extension of L_0 by L_2 . In this section we shall work in the language L_2 and in an extension $\hat{L}_2$ of L_2 .

Now we recall the basic system of open axioms for the extended language. It is called simply *BASIC* and it plays a similar role to Q in the usual language.

4.1 Definition. *BASIC* is the following theory:

- (1) $y \leq x \rightarrow y \leq Sx$;
- (2) $x \neq Sx$;
- (3) $\bar{0} \leq x$;
- (4) $x \leq y \ \& \ x \neq y \equiv Sx \leq y$;
- (5) $x \neq \bar{0} \rightarrow \bar{2}x \neq \bar{0}$;
- (6) $y \leq x \vee x \leq y$;
- (7) $x \leq y \ \& \ y \leq x \rightarrow x = y$;
- (8) $x \leq y \ \& \ y \leq z \rightarrow x \leq z$;
- (9) $|\bar{0}| = \bar{0}$;
- (10) $x \neq \bar{0} \rightarrow |\bar{2}x| = S(|x|) \ \& \ |S(\bar{2}x)| = S(|x|)$;
- (11) $|\bar{1}| = \bar{1}$;
- (12) $x \leq y \rightarrow |x| \leq |y|$;
- (13) $|x \# y| = S(|x| * |y|)$;
- (14) $\bar{0} \# y = \bar{1}$;
- (15) $x \neq \bar{0} \rightarrow 1 \# (\bar{2}x) = \bar{2}(1 \# x) \ \& \ 1 \# (S(\bar{2}x)) = \bar{2}(1 \# x)$;
- (16) $x \# y = y \# x$;
- (17) $|x| = |y| \rightarrow x \# z = y \# z$;
- (18) $|x| = |u| + |v| \rightarrow x \# y = (u \# y) * (v \# y)$;
- (19) $x \leq x + y$;
- (20) $x \leq y \ \& \ x \neq y \rightarrow S(\bar{2}x) \leq \bar{2}y \ \& \ S(\bar{2}x) \neq \bar{2}y$;
- (21) $x + y = y + x$;
- (22) $x + \bar{0} = x$;
- (23) $x + Sy = S(x + y)$;

- (24) $(x + y) + z = x + (y + z)$;
 (25) $x + y \leq x + z \equiv y \leq z$;
 (26) $x * \bar{0} = \bar{0}$;
 (27) $x * (Sy) = xy + x$;
 (28) $xy = yx$;
 (29) $x(y + z) = xy + xz$;
 (30) $x \geq \bar{1} \rightarrow (xy \leq xz \equiv y \leq z)$;
 (31) $x \neq \bar{0} \rightarrow |x| = S(\lfloor x/2 \rfloor)$;
 (32) $y = \lfloor x/2 \rfloor \equiv (\bar{2}y = x \vee S(\bar{2}y) = x)$,

where $\bar{1}$ and $\bar{2}$ are numerals as defined in Sect. 3.

This system is probably not quite optimal. One can play with it in order to find a shorter or nicer system. For instance, if we redefine

$$|x| = \lceil \log_2(x+1) \rceil - 1 \quad \text{if } x > 0, \\ 0 \quad \text{otherwise,}$$

then $x \# y = 2^{|x|*|y|}$ becomes associative and distributive with respect to multiplication, hence one would obtain a more homogeneous axiomatic system for such functions. We shall not do it here, as we want to concentrate on more interesting questions.

We shall use the bounded quantifiers $(\forall x \leq \tau)$ and $(\exists x \leq \tau)$; note that τ is a term in the language L_2 , not just in L_0 (not containing x , of course). Moreover we also need the so-called *sharply bounded quantifiers*. They have the form

$$(\forall x \leq |\tau|) \text{ and } (\exists x \leq |\tau|),$$

i.e. the outermost function in the bounding term is the length function. Their relation to ordinary bounded quantifiers in bounded arithmetic is similar to the relation of bounded quantifiers to unbounded quantifiers in *PA*. The classes Σ_i^b and Π_i^b of bounded formulae are defined as follows.

4.2 Definition.

- (1) $\Pi_0^b = \Sigma_0^b$ consist of formulae with sharply bounded quantifiers only;
- (2) Σ_{i+1}^b and Π_{i+1}^b are the least sets satisfying
 - (a) $\Sigma_i^b, \Pi_i^b \subseteq \Sigma_{i+1}^b$, and $\Sigma_i^b, \Pi_i^b \subseteq \Pi_{i+1}^b$;
 - (b) $\alpha \in \Sigma_{i+1}^b \Rightarrow (\exists x \leq \tau) \alpha \in \Sigma_{i+1}^b, (\forall x \leq |\tau|) \alpha \in \Sigma_{i+1}^b$,
 $\alpha \in \Pi_{i+1}^b \Rightarrow (\forall x \leq \tau) \alpha \in \Pi_{i+1}^b, (\exists x \leq |\tau|) \alpha \in \Pi_{i+1}^b$;
 - (c) $\alpha, \beta \in \Sigma_{i+1}^b \Rightarrow \alpha \& \beta, \alpha \vee \beta \in \Sigma_{i+1}^b$,
 $\alpha, \beta \in \Pi_{i+1}^b \Rightarrow \alpha \& \beta, \alpha \vee \beta \in \Pi_{i+1}^b$;
 - (d) $\alpha \in \Sigma_{i+1}^b, \beta \in \Pi_{i+1}^b \Rightarrow \neg \beta, \beta \rightarrow \alpha \in \Sigma_{i+1}^b$,
 $\alpha \in \Pi_{i+1}^b, \beta \in \Sigma_{i+1}^b \Rightarrow \neg \beta, \beta \rightarrow \alpha \in \Pi_{i+1}^b$.

- (3) α is in Δ_i^b with respect to a theory T , if α is equivalent to a Σ_i^b formula and a Π_i^b formula in T .
- (4) $\Sigma_0^b(\Sigma_i^b)$ is defined as the closure of Σ_i^b under connectives and sharply bounded quantification; note that this class is sometimes denoted also by $\Sigma_{i+1}^b \cap \Pi_{i+1}^b$.

Later on we shall see that, for $i \geq 1$, Σ_i^b (resp. Π_i^b) formulae define just the sets in Σ_i^p (resp. Π_i^p). The most important formalizations of induction for bounded formulae are the usual schema of induction and the following ones.

4.3 Definition.

- (1) $PIND(\alpha(x))$ is the formula

$$\alpha(\bar{0}) \& (\forall x)(\alpha(\lfloor x/2 \rfloor) \rightarrow \alpha(x)). \rightarrow (\forall x)\alpha(x);$$

- (2) $LIND(\alpha(x))$ is the formula

$$\alpha(\bar{0}) \& (\forall y < |x|)(\alpha(y) \rightarrow \alpha(S(y))). \rightarrow \alpha(|x|);$$

- (3) $LMIN(\alpha(x))$ is the formula

$$(\exists x)\alpha(x) \rightarrow \alpha(\bar{0}) \vee (\exists x)(\alpha(x) \& (\forall y \leq \lfloor x/2 \rfloor)(\neg \alpha(y)));$$

- (4) $PIND\Gamma$ ($LIND\Gamma$, $LMIN\Gamma$, respectively) is the schema (or the set) $PIND(\alpha(x))$ ($LIND(\alpha(x))$, $LMIN(\alpha(x))$, respectively) for $\alpha(x) \in \Gamma$.

Here we assume that $\alpha(x)$ may contain other free variables, which are called parameters. We have written $LIND$ in the form which contains only one additional quantifier and this quantifier is sharply bounded. Thus if α is Δ_i^b , then $LIND(\alpha(x))$ is also Δ_i^b ; if $\alpha(x)$ is in $\Pi_i^b \cup \Sigma_i^b$, then $LIND(\alpha(x))$ is $\Sigma_0^b(\Sigma_i^b)$.

Now we are ready to define the hierarchy of subtheories of bounded arithmetic.

4.4 Definition.

- (1) For $i \geq 0$, S_2^i is $BASIC + PIND\Sigma_i^b$.
- (2) For $i \geq 0$, T_2^i is $BASIC + I\Sigma_i^b$.
- (3) S_2 is $\bigcup_i S_2^i$; T_2 is $\bigcup_i T_2^i$.

The most important fragment of S_2 is S_2^1 . It serves as a basic theory modulo which results on fragments of S_2 are proved. Its role is similar to

the role of $I\Sigma_1$ in PA . In particular we shall show in the next subsection that all polynomial time computable functions are definable and provably total in S_2^1 . As for weaker fragments: S_2^0 is too weak (it does not prove the existence of the predecessor, see [Takeuti, Sharply Bounded]); the strength of T_2^0 is not quite known. We shall not repeat the formalization of elementary mathematical concepts in S_2^1 which has been done in [Buss 86, Bounded Arith.]. One can use some techniques from Sect. 3, mainly those which were mentioned in connection with $I\Sigma_0 + \Omega_1$. (Those which reduced the size of the quantified numbers by using more quantifiers are useless for S_2^1 , since we need Σ_1^b definitions.) Here we briefly mention a definition of coding in S_2^1 .

First we re-define $Pow(x)$ (x is a power of 2) by an open formula in L_2 :

$$Pow(x) \equiv \bar{2}x = x \# \bar{1}.$$

In order to see that this works, we shall show in *BASIC* that Pow satisfies the following recursive condition:

$$Pow(x) \rightarrow (x = \bar{1} \vee (\exists y \leq x)(x = \bar{2}y \ \& \ Pow(y))).$$

One can easily show $\neg Pow(\bar{0})$. Suppose $x > 1$. Then, by axiom (32) of *BASIC*, $x = \bar{2}y$ or $x = S(\bar{2}y)$, for $y = \lfloor x/2 \rfloor$ and $y > 0$. By the definition of Pow and (15), we have $\bar{2}x = x \# \bar{1} = \bar{2}(y \# \bar{1})$, hence (by (1), (6), (7) and (30)) we have $x = y \# \bar{1}$. Since $y > 0$, we get from (15) that x is even. Hence we must have the first possibility $x = \bar{2}y$ (*BASIC* proves that $S(\bar{2}u) \neq \bar{2}v$). Thus $\bar{2}y = x = y \# \bar{1}$, i.e. $Pow(y)$.

Now we can re-define $bit(x, i)$ (the i -th digit of x) by a Σ_1^b formula:

$$\begin{aligned} bit(x, i) = z &\equiv z \leq 1 \ \& \ (\exists y, u_2 \leq x)(\exists u_1 < y)(Pow(y) \ \& \ |y| = i \\ &\quad \& \ x = \bar{2}u_2y + zy + u_1), \text{ if } i > 0, \\ &= \bar{0} \text{ otherwise.} \end{aligned}$$

Since the decomposition above is unique (see Theorem I.1.30), we have also a Π_1^b definition:

$$\begin{aligned} bit(x, i) = z &\equiv (\forall w \leq 1)(\forall y, u_2 \leq x)(\forall u_1 < y)(Pow(y) \ \& \ |y| = i \\ &\quad \& \ x = \bar{2}u_2y + wy + u_1 \rightarrow w = z), \text{ if } i > 0, \\ &= \bar{0} \text{ otherwise.} \end{aligned}$$

Thus we have a Δ_1^b definition of bit in S_2^1 . Similarly we can define function sgm , (see 3.18). The function bit codes 0-1 sequences. If we want to code arbitrary sequences, we can proceed as follows. A sequence will be a triple $p = (u, v, w)$ (more precisely $((u, v), w)$, thus every number is a sequence),

where u is the code of the sequence, v is its *maximal element* and w is the *length of the sequence*, i.e.

$$\begin{aligned}\max(p) &= v, \\ lh(p) &= w.\end{aligned}$$

The function the i -th element of p is defined by

$$\begin{aligned}(p)_i &= x \equiv x \leq v \ \& \ (\forall j \leq |v|)(bit(x, j) = bit(u, iv + j)), \text{ if } i < w, \\ (p)_i &= \bar{0} \text{ otherwise.}\end{aligned}$$

Hence all these functions are Δ_1^b . Finally we observe that we can bound u by

$$|u| \leq |v|.w.$$

Thus for some term *bound* we have

$$(a.1) \quad p \leq bound(\max(p), 2^{lh(p)}).$$

4.5 Proposition.

- (1) For $i \geq 1$, $S_2^i \equiv S_2^1 + PIND \Pi_i^b \equiv S_2^1 + LIND \Sigma_i^b \equiv S_2^1 + LIND \Pi_i^b$.
- (2) For $i \geq 1$, $T_2^i \equiv S_2^1 + III_i^b$.

Proof. We shall show only $S_2^i \equiv S_2^1 + LIND \Sigma_i^b$; the proofs of the remaining equivalences are the same as for fragments of PA , see Theorem I.2.4.

First we shall derive $LIND \Sigma_i^b$ in S_2^1 . Let $\varphi(x)$ be Σ_i^b . Let $\psi(x)$ be $\varphi(|x|)$. Then by (10) and (32) of *BASIC*, the inductive clause of $LIND(\varphi(x))$

$$\varphi(x) \rightarrow \varphi(S(x))$$

implies the inductive clause of $PIND(\psi(x))$

$$\psi(\lfloor x/2 \rfloor) \rightarrow \psi(x).$$

Thus $LIND \Sigma_i^b$ is reduced to $PIND \Sigma_i^b$.

Now we shall derive $PIND \Sigma_i^b$ in $S_2^1 + LIND \Sigma_i^b$. Let $\varphi(x)$ be a Σ_i^b formula for which we want to prove $PIND$. Thus we assume $\varphi(\bar{0})$ and $(\forall x)(\varphi(\lfloor x/2 \rfloor) \rightarrow \varphi(x))$. Let $msp(x, y)$ be the number whose binary representation consists of the first y digits of x (and, say, it is x if $y > |x|$). This function has a Δ_1^b definition in S_2^1 (use the function sgm). The idea of the proof is to show, using $LIND$, that all end segments of x satisfy φ , hence also x itself satisfies φ . Formally, we apply $LIND$ to $\varphi(msp(x, y))$ with y as the induction variable and x as a parameter. Thus $PIND(\varphi(x))$ is reduced to

$LIND(\varphi(msp(x, y)))$ since $msp(x, 0) = 0$, $\lfloor msp(x, S(y))/2 \rfloor = msp(x, y)$ and $msp(x, |x|) = x$. $\square$

4.6 Theorem. For $i \geq 1$, $S_2^i \vdash I\Delta_1^b$.

Proof. Recall the meaning of $I\Delta_1^b$: for every $\alpha \in \Sigma_1^b$ and $\beta \in \Pi_1^b$

$$(\forall x)(\alpha(x) \equiv \beta(x)) \rightarrow I(\alpha(x)).$$

The idea of the proof is the following. The schema of induction tells us that we cannot define a proper subset of numbers which contains 0 and is closed under the successor. The schema $PIND$ implies that there is no subset of numbers containing 0 and closed under the successor *and the function* $2x$. We already know (Theorem II.3.5) that for every cut we can define a subcut closed under addition, hence also closed under $2x$. We shall use this construction to reduce induction to $PIND$. There is a small technical complication, since we have to use only bounded formulae, while the original construction uses unbounded quantifiers.

Let α and β be given. Put

$$\psi(y, a) \equiv_{df} (\forall x \leq a)(\alpha(x) \rightarrow \beta(\min(x + y, a))).$$

Thus ψ is Π_1^b . Working in S_2^i , assume $\alpha(x) \equiv \beta(x)$, $\alpha(0)$ and $(\forall x)(\alpha(x) \rightarrow \alpha(S(x)))$. Furthermore assume that $\alpha(a)$ fails for some a . We have $\psi(0, a)$ trivially and $\psi(y, a) \rightarrow \psi(S(y), a)$, since $(\forall x)(\alpha(x) \rightarrow \alpha(S(x)))$. We shall prove $\psi(y, a) \rightarrow \psi(2y, a)$. Assume $\psi(y, a)$ and $\alpha(x)$ for some $x \leq a$. By $\psi(y, a)$ we have $\alpha(\min(x + y, a))$. Thus $x + y < a$, since $\neg\alpha(a)$. Hence $\alpha(x + y, a)$. Applying $\psi(y, a)$ once again (with $x + y$ instead of x) we get $\alpha(\min(x + 2y, a))$. Hence we have $\psi(y, a) \rightarrow \psi(2y, a)$. Since

$$x = 2\lfloor x/2 \rfloor \vee x = S(2\lfloor x/2 \rfloor),$$

the formula

$$(\forall y)(\psi(y, a) \rightarrow \psi(S(y), a) \ \& \ \psi(\bar{2}y, a))$$

implies

$$(\forall y)(\psi(\lfloor y/2 \rfloor, a) \rightarrow \psi(y, a)).$$

Hence using $PIND(\psi)$ we get $\psi(a, a)$. Since $\alpha(0)$, this implies $\alpha(a)$. We have derived $\alpha(a)$ from its negation, hence it must be true. Thus we have shown induction for α . $\square$

Note that, for $i \geq 2$, a weaker version of this theorem can be derived also from Corollary 4.28 below using the Σ_{i+1}^b conservativity of S_2^{i+1} over T_2^i (Corollary 4.34).

4.7 Theorem.

- (1) For $i \geq 0$, $S_2^i \subseteq T_2^i \subseteq S_2^{i+1}$;
 (2) $S_2 \equiv T_2$ and they are conservative extensions of $I\Sigma_0 + \Omega_1$.

Proof. (1) To prove the first inclusion, observe that T_2^i trivially proves $LIND \Sigma_i^b$ and apply Theorem 4.5 (1). The second inclusion is a corollary of Theorem 4.6.

(2) The equivalence of theories S_2 and T_2 follows from (1) of the preceding proposition. In order to show that $I\Sigma_0 + \Omega_1$ is contained in S_2 we have only to prove axiom Ω_1 in S_2 , which follows easily from the bound $\omega_2(x) \leq (x \# x)^4$.

Let us prove the conservativity of S_2 over $I\Sigma_0 + \Omega_1$. We already know that $\lfloor x/2 \rfloor$ and $|x|$ are definable in $I\Sigma_0 + \Omega_1$. As we have a Σ_0 definition of the exponentiation relation, we can define the graph of the function $x \# y$. This function is provably total in $I\Sigma_0 + \Omega_1$, because of the bound $x \# y \leq \omega_2(x+y)$. The axioms of *BASIC* are easy too. To prove induction for bounded formulae in language L_2 in $I\Sigma_0 + \Omega_1$, recall that by Proposition 1.3 we can eliminate terms as bounds in the formulae used in induction. Then we use the definitions of the new operations to translate the formulae into the language L_0 and apply $I\Sigma_0$. $\square$

4.8 Theorem. (a) For $i \geq 1$, and $\alpha(u, x) \in \Sigma_0^b(\Sigma_i^b)$,

$$S_2^i \vdash (\exists w)(\forall x < |y|)(\alpha(u, x) \equiv \text{bit}(w, x + \bar{1}) = \bar{1}).$$

(b) For $i \geq 1$, $S_2^i \vdash LIND(\Sigma_0^b(\Sigma_i^b))$.

Note that (a) is a version of the $\Sigma_0^b(\Sigma_i^b)$ comprehension schema in S_2^i where we use $\text{bit}(w, x)$ as a convenient coding of 0-1 sequences.

Proof. (a) Consider a more general schema for $\alpha \in \Sigma_0^b(\Sigma_i^b)$:

$$(b.1) \quad S_2^i \vdash (\exists w)(\forall x_1 < |y_1|) \dots (\forall x_n < |y_n|)(\alpha(u, x_1, \dots, x_n) \equiv \text{bit}(w, \bar{1} + (x_1, (x_2, \dots, x_n))) = \bar{1}).$$

We should show that (b.1) holds for $\alpha \in \Sigma_i^b$ and that the set of formulae α satisfying (b.1) is closed under boolean operations and sharply bounded quantifications.

Suppose $\alpha \in \Sigma_i^b$. Take the formula $\beta(z, u)$ defined by

$$\begin{aligned} & (\exists w)(|w| \leq |y_1| * |y_2| * \dots * |y_n| \ \& \ \text{nuon}(w) = z \\ & \ \& \ (\forall x_1 < |y_1|) \dots (\forall x_n < |y_n|)(\text{bit}(w, \bar{1} + (x_1, (x_2, \dots, x_n))) = \bar{1} \rightarrow \\ & \quad \rightarrow \alpha(u, x_1, \dots, x_n))). \end{aligned}$$

Let us work in S_2^i . This formula is true for $z = 0$ and fails for $z = |y_1| * |y_2| * \dots * |y_n| + 1$, since $\text{nuon}(w) \leq |w|$. Since α is Σ_i^b , nuon is Δ_1^b and we can bound w , say by $y_1 \# y_2 \# \dots \# y_n$, the formula β is also Σ_i^b . Hence we can use $LIND \Pi_i^b$ to find the smallest z for which $\beta(z, u)$ fails. Then take w such that $\beta(z - 1, w)$. If

$$z - 1 = |y_1| * |y_2| * \dots * |y_n|,$$

then all bits of w are 1, hence we have (b.1) for this w . Now suppose that

$$z - 1 < |y_1| * |y_2| * \dots * |y_n|.$$

If (b.1) were not true for this w , then we could add another 1 in the binary representation of w which codes an instance of $\alpha(u, x_1, x_2, \dots, x_n)$. But this would contradict the condition that $z - 1$ is the last parameter for which $\beta(z - 1, w)$ is true. Hence w codes exactly the set determined by α .

The case of propositional connectives and sharply bounded quantifiers is easy.

(b) follows immediately from (a). $\square$

In Bounded Arithmetic there is a counterpart of the *collection axiom* for Peano Arithmetic, see I.2.1. We shall call it *the bounded collection axiom*. Sometimes it is also called *replacement axiom*.

4.9 Definition. (i) $BB(\alpha(x, y))$ *bounded collection* is the following formula

$$(\forall x \leq |t|)(\exists y \leq s)\alpha(x, y). \rightarrow (\exists w)(\forall x \leq |t|)((w)_x \leq s \ \& \ \alpha(x, (w)_x)).$$

(ii) $BB\Gamma$ is the schema $BB(\alpha)$ for $\alpha \in \Gamma$.

Using (a.1) above, we can bound the variable w in the bounded collection axiom; thus it has the following form (equivalent in S_2^1):

$$\begin{aligned} &(\forall x \leq |t|)(\exists y \leq s)\alpha(x, y). \rightarrow \\ &\rightarrow (\exists w \leq \tau(s, t))(\forall x \leq |t|)((w)_x \leq s \ \& \ \alpha(x, (w)_x)), \end{aligned}$$

where τ is a suitable term ($\tau(s, t) = \text{bound}(s, 2^{|t|+1})$).

4.10 Theorem. For $i \geq 1$, $S_2^i \vdash BB \Sigma_i^b$.

Proof. Let $\alpha(x, y)$ be a Σ_i^b formula. Assume

$$(\forall x \leq |t|)(\exists y \leq s)\alpha(x, y).$$

Then we obtain

$$(\exists w \leq \tau(s, t))(\forall x \leq |t|)((w)_x \leq s \ \& \ \alpha(x, (w)_x))$$

by

$$LIND((\exists w \leq \tau(s, t))(\forall x \leq |z|)((w)_x \leq s \& \alpha(x, (w)_x))),$$

where the induction variable is z . The formula is Σ_i^b , since coding is Δ_1^b . We use the elementary facts that the empty sequence has a code, each sequence can be extended by an arbitrary element and the estimate of the size of a sequence. $\square$

Note that the implication converse to that above is already provable in S_2^1 (using only the fact that $(w)_x$ is a definable function in S_2^1). The bounded form of the bounded collection axiom enables us to interchange sharply bounded quantifiers with bounded quantifiers. However we cannot push inside all sharply bounded quantifiers without extending the language, since the coding function is not defined by a term in L_2 . Therefore we need to extend the language.

4.11 Definition. (i) The language L_2 augmented with the binary operation $(x)_z$, will be denoted by $\hat{L}_2$; the corresponding classes Σ_i^b, Π_i^b etc. will be denoted by $\hat{\Sigma}_i^b, \hat{\Pi}_i^b$, etc.

(ii) We say that a formula is *strict* $\hat{\Sigma}_i^b$ if it is of the form

$$(\exists x_1 \leq t_1)(\forall x_2 \leq t_2) \dots (Q x_i \leq t_i) \varphi,$$

where Q is $\exists$ if i is odd and $\forall$ if i is even, and φ is sharply bounded.

We shall not introduce new notation for theories obtained by extending the language and adding the definition of the coding relation. It will always be clear from the context which theory we mean.

4.12 Corollary. For $i \geq 1$, every Σ_i^b formula is equivalent to a *strict* $\hat{\Sigma}_i^b$ formula, provably in S_2^i (augmented with the definition of $(x)_z$).

Proof. We can shift all sharply bounded quantifiers beyond bounded (but not sharply) quantifiers using the bounded collection schema. Successive quantifiers of the same kind can be merged using the coding function. $\square$

4.13 Corollary. For $i \geq 1$ and $\alpha(x, y)$ in Σ_i^b ,

$$S_2^i \vdash (\exists w)(\forall x \leq |t|)[(\exists y \leq s)\alpha(x, y) \equiv ((w)_x \leq s \& \alpha(x, (w)_x))].$$

(This is sometimes called *strong replacement*.)

Proof. Let $\alpha(x, y)$ in Σ_i^b be given. By Theorem 4.8, we can code the set of x 's such that $(\exists y \leq s)\alpha(x, y)$ using some v and formula $\text{bit}(v, x) = 1$. Now take $\alpha'(x, y)$ to be

$$\alpha(x, y) \vee \text{bit}(v, x) = 0.$$

Then we have $(\forall x \leq |t|)(\exists y \leq s)\alpha'(x, y)$, hence we can apply bounded collection; the sequence w given by it satisfies the clause of the theorem. $\square$

(c) Definability of Turing Machine Computations in Fragments of Bounded Arithmetic

Recall that for $i \geq 1$, $\square_i^p$ denotes the set of functions computable in polynomial time using oracles for Σ_{i-1}^p sets; in particular, $\square_1^p$ are just the polynomial time computable functions, since Σ_0^p oracles are superfluous. In this subsection we shall show that functions from $\square_1^p$ have suitable definitions in S_2^1 , and the same holds for $\square_i^p$ and T_2^{i-1} , if $i > 1$. This means that for functions from $\square_i^p$ we have Δ_{i+1}^b definitions in the corresponding theories and that basic properties of the functions are provable in these theories. We shall only show that the defining formulae have appropriate complexity and that they define total functions. Other properties follow from the construction and their proofs are omitted.

We shall consider deterministic Turing machines, possibly with oracles. Let an oracle Turing machine M be given, let e be the code of it, let A be an oracle. We formalize the computation of M on input a as two sequences w and q , where $(w)_i$ is the instantaneous description of the state of M , the position of the heads and the content of the tapes, and $(q)_i$ is the query asked in step i .¹ The computation is determined by a constant $firstq$ (let us set it equal to 0) and five functions $firstw$, $nextw_0$, $nextw_1$, $nextq_0$, $nextq_1$ as follows:

$$\begin{aligned}(w)_0 &= firstw(a), & (q)_0 &= firstq; \\ (w)_{i+1} &= nextw_0((w)_i, (q)_i), & (q)_{i+1} &= nextq_0((w)_i, (q)_i), & \text{if } (q)_i \notin A; \\ (w)_{i+1} &= nextw_1((w)_i, (q)_i), & (q)_{i+1} &= nextq_1((w)_i, (q)_i), & \text{if } (q)_i \in A.\end{aligned}$$

We can think of $(w)_i$ as a sequence which encodes the current situation on the tapes and encodes the state of M . Then the functions above are only local transformations of these sequences; hence it is not difficult to write down Δ_1^b definitions for them using our Δ_1^b definition of coding. If the oracle A is defined by a formula $\varphi(x)$, we thus get a formula

$$Comp_{M, \varphi(x)}(w, q, a)$$

expressing that (w, q) is a computation of M on input a . If M is a machine without an oracle, we shall write simply $Comp_M(w, a)$. We shall assume that

¹ W.l.o.g. we can assume that M asks a query in each computation step; if it does not need information it asks some default fixed query.

$(w)_i = (w)_{i+1}$ iff M stops in the i -th step. Thus computations of length n exist for every n , and we have a simple condition for testing the termination. Furthermore we shall assume that *this* $(w)_i$ is the output of M . We shall identify oracles with their defining formulae; thus we shall talk about Σ_i^b oracles instead of Σ_i^P oracles.

4.14 Lemma.

- (i) $\text{Comp}_M(w, a)$ is Δ_1^b in S_2^1 .
- (ii) If $\varphi(x)$ is Σ_i^b , $i \geq 1$, then $\text{Comp}_{M, \varphi(x)}(w, q, a)$ is $\Sigma_0^b(\Sigma_i^b)$.
- (iii) There exists a term σ such that

$$S_2^1 \vdash \text{Comp}_{M, \varphi(x)}(w, q, a) \rightarrow w, q \leq \sigma(a, lh(w)).$$

Proof. (i) and (ii). We have

$$\text{Comp}_M(w, a) \equiv (w)_0 = \text{firstw}(a) \ \& \ (\forall i < lh(w) - 1)((w)_{i+1} = \text{nextw}((w)_i)),$$

and

$$\begin{aligned} \text{Comp}_{M, \varphi(x)}(w, q, a) \equiv & (w)_0 = \text{firstw}(a) \ \& \ (\forall i < lh(w) - 1) \\ & [(\neg \varphi(q)_i \ \& \ (w)_{i+1} = \text{nextw}_0((w)_i, (q)_i) \ \& \ (q)_{i+1} = \text{nextq}_0((w)_i, (q)_i)) \\ & \vee (\varphi(q)_i \ \& \ (w)_{i+1} = \text{nextw}_1((w)_i, (q)_i) \ \& \ (q)_{i+1} = \text{nextq}_1((w)_i, (q)_i))]. \end{aligned}$$

We already know that all the functions used in these formulae are Δ_1^b definable.

(iii) The length of w and q is polynomial in the number of steps of the computation and the length of input, whence we get the bound. $\square$

Now we prove the relation between the classes of formulae Σ_i^b and the classes of sets Σ_i^P which we have promised. Note, however, that the relation is not quite direct for $i = 0$.

4.15 Theorem.

- (i) Every set definable by a Σ_0^b formula is in *LogSpace*.
 - (ii) Every set in P is definable by a Δ_1^b formula.
 - (iii) For $i \geq 1$, the sets in Σ_i^P (resp. Π_i^P) are just the sets definable by Σ_i^b (resp. Π_i^b) formulae.
- (“Definable” means here “definable in the standard model”).

Proof. (i) All the functions and predicates of the language of S_2 are *LogSpace* computable and sharply bounded quantification preserves *LogSpace* computations.

(ii) Let X be in P . Let M be a Turing machine which always stops after $p(|a|)$ steps on input a and outputs 0 iff a belongs to X . Then we can define X by

$$(\exists w)(lh(w) \leq p(|a|) \& \text{Comp}_M(w, a) \& (w)_{p(|a|)} = 0),$$

or by

$$(\forall w)(lh(w) \leq p(|a|) \& \text{Comp}_M(w, a) \rightarrow (w)_{p(|a|)} = 0).$$

By Lemma 4.14, the inner parts of the formulae are Δ_1^b ; by the same lemma the bound $lh(w) \leq p(|a|)$ implies that w can be bounded by a term in a , hence the formulae are equivalent to Σ_1^b and Π_1^b formulae respectively, thus they are Δ_1^b .

(iii) By Corollary 4.12 every Σ_i^b formula is equivalent to a strict Σ_i^b formula in the language extended by the coding relation; thus it can be written in the form

$$(\exists y_1 \leq t_1)(\forall y_2 \leq t_2) \dots (Qy_i \leq t_i) \Phi,$$

where Φ is sharply bounded in the extended language and where Q is $\exists$ if i is odd and $\forall$ if i is even. Since the coding relation also belongs to P , formula Φ defines a predicate in P . The bounds are just of the order needed for the definition of Σ_i^p and the relations $y_j \leq t_j$ are in P as well. Thus, by Theorem 2.11, each set definable by a Σ_i^b formula is in Σ_i^p .

To prove the converse let X be in Σ_i^p . Thus X is defined by

$$(\exists y_1, |y_1| \leq p_1(|x|)) \dots (Qy_i, |y_i| \leq p_i(|x|)) P(x, y_1, \dots, y_i),$$

where P is in P . By (ii) we can define P by a Δ_1^b formula. Similarly to above we can replace the bounds $|y_j| \leq p_j(|x|)$ by bounds of the form $y_j \leq t_j$ for suitable terms $t_1, \dots, t_i$. Thus we obtain a Σ_i^b formula defining X . $\square$

4.16 Theorem. (1) Let M be a Turing machine. Then S_2^1 proves that for every input a and every b , there exists a unique computation of M which has $|b|$ steps. Formally:

$$(i) \quad S_2^1 \vdash (\exists w)(\text{Comp}_M(w, a) \& lh(w) = |b|);$$

$$(ii) \quad S_2^1 \vdash \text{Comp}_M(w, a) \& \text{Comp}_M(w', a) \& lh(w), lh(w') = |b| \rightarrow w = w'.$$

(2) Let $i \geq 1$, let M be an oracle Turing machine and let $\varphi(x)$ be in Σ_i^b . Then T_2^i proves that for every input a and every b , there exists a unique computation of M with oracle $\varphi(x)$ which has $|b|$ steps. Formally:

$$(i) \quad T_2^i \vdash (\exists w, q)(\text{Comp}_{M, \varphi(x)}(w, q, a) \& lh(w) = |b|);$$

$$(ii) \quad T_2^i \vdash \text{Comp}_{M, \varphi(x)}(w, q, a) \& \text{Comp}_{M, \varphi(x)}(w', q', a) \& lh(w), lh(w') = |b| \rightarrow w = w' \& q = q'.$$

Proof. Let M and φ be given, φ in Σ_i^b . Let ψ be defined by

$$\begin{aligned}\psi(w, q, a, b) \equiv & lh(w) = |b| + 1 \ \& \ (w)_0 = firstw(a) \ \& \ (q)_0 = firstq \ \& \ (\forall j < |b|) \\ & [((w)_{j+1} = nextw_0((w)_j, (q)_j) \ \& \ (q)_{j+1} = nextq_0((w)_j, (q)_j)) \\ & \vee (\varphi((q)_j) \ \& \ (w)_{j+1} = nextw_1((w)_j, (q)_j) \\ & \ \& \ (q)_{j+1} = nextq_1((w)_j, (q)_j))].\end{aligned}$$

The meaning of the formula is that w, q is like a computation in which we follow the advice of the oracle if its answer is negative, but not always do we follow it if its answer is positive. This formula is Δ_1^b if $i = 0$, and Σ_i^b if $i > 0$. Hence we can prove, using respectively $LIND \Sigma_1^b$ or $LIND \Sigma_i^b$ on b , that for given a, b there exists at least one pair w, q such that $\psi(w, q, a, b)$. Using $I\Sigma_i^b$ we now minimize the number of steps where w and q do not follow the positive advice. Let Ψ be the Σ_i^b formula defined by

$$\begin{aligned}\Psi(u, a, b) \equiv & (\exists w, q \leq \sigma(a, |b|))(\psi(w, q, a, b) \ \& \ (\forall j < |b|) \\ & [(w)_{j+1} = nextw_0((w)_j, (q)_j) \ \& \ (q)_{j+1} = nextq_0((w)_j, (q)_j) \rightarrow \\ & \rightarrow bit(u, |b| - j) = 1]),\end{aligned}$$

where σ is the term from Lemma 4.14 (iii). For a and b there exists at least one u satisfying $\Psi(u, a, b)$, since for w, q satisfying $\psi(w, q, a, b)$ the existence of u follows by $LIND \Delta_1^b$. By $I\Sigma_i^b$ we have the least u for which $\Psi(u, a, b)$ holds (this is the place where we use the full strength of T). Let w, q be the witnesses of $\Psi(u, a, b)$ for this u . We shall show that for this w, q also the implication complementary to the one in ψ holds for all $j < |b|$:

$$\varphi((q)_j) \rightarrow (w)_{j+1} = nextw_1((w)_j, (q)_j) \ \& \ (q)_{j+1} = nextq_1((w)_j, (q)_j).$$

Suppose not, then we have $\varphi((q)_j)$ and

$$(w)_{j+1} = nextw_0((w)_j, (q)_j) \ \& \ (q)_{j+1} = nextq_0((w)_j, (q)_j),$$

for some $j < |b|$. Hence $bit(u, |b| - j) = 1$. Take w' and q' of length $|b|$ such that

$$(w')_k = (w)_k \ \& \ (q')_k = (q)_k,$$

for $k = 0, \dots, j$,

$$(w')_{j+1} = nextw_1((w)_j, (q)_j) \ \& \ (q')_{j+1} = nextq_1((w)_j, (q)_j),$$

and

$$(w')_{k+1} = nextw_0((w')_k, (q')_k) \ \& \ (q')_{k+1} = nextq_0((w')_k, (q')_k),$$

for $k = j + 1, \dots, |b| - 2$. The existence of such w' and q' is again easily provable by $LIND \Sigma_1^b$. Let u' be determined by

$$\begin{aligned} bit(u', k) &= 1, \quad \text{for } k = 1, \dots, |b| - j - 1, \\ bit(u', |b| - j) &= 0, \\ bit(u', k) &= bit(u, k), \quad \text{for } k = |b| - j + 1, \dots, |b|. \end{aligned}$$

Then we have $\Psi(u', a, b)$, since u' corresponds to w', q' , but $u' < u$, which is a contradiction. Thus we have shown

$$Comp_{T, \varphi(x)}(w, q, a) \& lh(w) = |b|.$$

To prove uniqueness we apply $LIND \Delta_1^b$ to

$$(\forall j < |b|)((w)_j = (w')_j \& (q)_j = (q')_j). \quad \square$$

4.17 Theorem. $\square_1^p$ functions are definable by formulae which are Δ_1^b in S_2^1 ; for $i > 0$, $\square_{i+1}^p$ functions are definable by formulae which are Δ_{i+1}^b in T_2^i . Furthermore, for each such formula the corresponding theory proves that the formula defines a total function.

Proof. Let $i \geq 1$. Let f be a $\square_{i+1}^p$ function defined by an oracle machine M and a Σ_i^p oracle A . By Theorem 4.15, A can be defined by a Σ_i^b formula $\varphi(x)$. Let $p(x)$ be a polynomial time bound for M . Then (similarly as in Theorem 4.15) $f(a) = b$ can be defined by

$$(\exists w, q \leq \sigma(a, p(|a|)))(Comp_{M, \varphi(x)}(w, q, a) \& (w)_{p(|a|)} = b),$$

or by

$$(\forall w, q \leq \sigma(a, p(|a|)))(Comp_{M, \varphi(x)}(w, q, a) \rightarrow (w)_{p(|a|)} = b),$$

where σ is the term from Lemma 4.14. Using the same lemma to estimate the complexity of $Comp_{M, \varphi(x)}$ and the existence and uniqueness of the computation (Theorem 4.16) we see that it is a Δ_{i+1}^b definition in T_2^i . For $i = 0$ the proof is similar. The fact that the formulae define total functions in the corresponding theories follows from Theorem 4.16. $\square$

In order to be able to use the definitions of functions for other constructions in T , we need more properties of formulae defining the functions. In particular, we shall use the following properties:

- (1) if $t(\mathbf{x})$ is a term and ψ is the definition of the corresponding $\square_1^p$ function, then $t(\mathbf{x}) = y$ is equivalent to $\psi(\mathbf{x}, y)$ in S_2^1 ;
- (2) $\square_i^p$ functions are provably closed under composition, i.e. if $\varphi(\mathbf{x}, y)$ defines $f(\mathbf{x})$ and $\psi_i(\mathbf{x}, y)$ define $g_i(\mathbf{x})$, then for some definition $\chi(\mathbf{x}, y)$ of $f(g_1(\mathbf{x}), \dots, g_n(\mathbf{x}))$ T_2^1 (resp. S_2^1 , if $i = 1$) proves

$$\chi(\mathbf{x}, y) \equiv (\exists z_1, \dots, z_n)(\psi_1(\mathbf{x}, z_1) \& \dots \& \psi_n(\mathbf{x}, z_n) \& \varphi(z_1, \dots, z_n, y));$$

- (3) $\square_i^p$ functions are closed under definitions by cases determined by Σ_{i-1}^b formulae (it is clear what is the corresponding formula);
- (4) $\square_i^p$ functions are closed under bounded recursion.

The last condition can be formalized as follows. Let $f(\mathbf{x}), g(z, \mathbf{x}), k(y, \mathbf{x})$ be functions in $\square_i^p$; then the function $h(y, \mathbf{x})$ defined by the following recursion must also belong to $\square_i^p$:

$$\begin{aligned} h(0, \mathbf{x}) &= f(\mathbf{x}), \\ h(y, \mathbf{x}) &= \min(g(h(\lfloor y/2 \rfloor, \mathbf{x}), \mathbf{x}), k(y, \mathbf{x})). \end{aligned}$$

We require that this schema be provable in T_2^1 (resp. S_2^1 , if $i = 1$) for the defining formulae of f, g and h . To prove these conditions one has only to check that the proofs in the standard model can be carried out in the fragments of Bounded Arithmetic. We omit the proofs since they are not difficult and contain no essentially new ideas.

So far we are able to talk about a single function form some class $\square_i^p$. Later we shall need formalization of Turing machine computations such that we can talk about Turing machines and oracles in the theory, i.e. we want a formula defining the computation which has parameters also for Turing machines and oracles. In the model theoretical language it means that we want to consider also nonstandard Turing machines and oracles.

A simple solution is to take a universal Turing machine, which is explicitly defined and thus has a formalization in S_2^1 by Theorem 4.17. Then the code of an arbitrary Turing machine will be the word which must be written on the input of the universal one in order to simulate it. A natural universal Turing machine simulates each machine M in such a way that the running time is at most polynomially longer than the running time of M . More precisely, the simulation time is bounded by a polynomial in the original running time and the size of M . In particular, if M runs in polynomial time so does the simulation of M .

This was about Turing machines. Oracles can be presented in a similar fashion. Using a universal Turing machine we can represent every predicate $P(x)$ in P in the form $\mu_0(e, x, t_e(x))$ for some fixed Δ_1^b formula μ_0 . Here e is the code (the index) of P and t_e is a suitable term determined by the running time of a Turing machine for P . This is still not quite what we need, since the dependence of t_e on e is not explicit. Thus we shall make an additional natural assumption that the predicate with code e is decidable in time n^e .

Then our representation of predicates in P will be of the form

$$\mu_0(e, x, \bar{2}^{|x|^e}).$$

Note that for e standard the existence of $\bar{2}^{|x|^e}$ is provable in S_2^1 . For the representation of predicates in Σ_i^b we have only to add quantifiers. Also there is no need to use two indices, one for the machine and one for the oracle. We shall present our conclusions as a theorem.

4.18 Theorem. There are formulae $\mu_0, \nu_0 \in \Delta_1^b$, $\mu_i \in \Sigma_i^b$, $\nu_i \in \Sigma_0^b(\Sigma_i^b)$ for $i \geq 1$, such that

- (i) each predicate which is in P , provably in S_2^1 , can be represented in S_2^1 as $\mu_0(\bar{e}, x, \bar{2}^{|x|^{\bar{e}}})$, for some numeral $\bar{e}$;
- (ii) for $i \geq 1$ and for every $\varphi(x) \in \Sigma_i^b$, there exists an e such that

$$S_2^1 \vdash \varphi(x) \equiv \mu_i(\bar{e}, x, \bar{2}^{|x|^{\bar{e}}});$$

- (iii) for every function f which is polynomial time computable, provably in S_2^1 , the relation $f(x) = y$ can be represented in S_2^1 by

$$\nu_0(\bar{e}, x, y, \bar{2}^{|x|^{\bar{e}}}), \quad \text{for some numeral } \bar{e};$$

- (iv) for $i \geq 1$ and for every function f which is in $\square_{i+1}^p$, provably in T_2^i , the relation $f(x) = y$ can be represented in T_2^i by

$$\nu_i(e, x, y, \bar{2}^{|x|^e}), \quad \text{for some numeral } \bar{e}.$$

Let us state (i) more precisely. A predicate P in P is determined by a 0-1 polynomial time computable function f . For this function we have a defining formula $\varphi(x, y)$ which is Δ_1^p and satisfies some basic conditions in S_2^1 . Thus $P(x)$ is defined by $\varphi(x, \bar{0})$. The precise statement of (i) is that for some number e

$$S_2^1 \vdash \varphi(x, \bar{0}) \equiv \mu_0(\bar{e}, x, \bar{2}^{|x|^{\bar{e}}}).$$

The statement (iii) should be understood in the same way.

In order to simplify notation we shall define

$$\{e\}_i(x) = y \equiv_{df} \nu_i(e, x, y, \bar{2}^{|x|^e}).$$

We should keep in mind that it is not a bounded formula. However, if we prove the existence of $\bar{2}^{|x|^e}$, then we can work with it as if it were a Δ_{i+1}^b formula.

(d) Witnessing Functions

In this subsection we shall develop the model theory of fragments of S_2 and prove some theorems of the type mentioned at the beginning of this section. We shall use the fact proved above that Turing machine computations are definable in fragments of S_2 .

First we shall show that a substructure of a model M of T_2^i that is closed under the functions of $\square_{i+1}^p$ is Σ_i^b elementary in M and is a model of T_2^i . The same holds for substructures of models of S_2^1 and the functions of $\square_1^p$. This is applied to prove witnessing Theorems 4.27 and 4.29 for fragments T_2^i . Then we prove a witnessing Theorem 4.32 for fragments S_2^{i+1} . The theorem has a stronger conclusion than in the original form used by Buss; it states that $(\forall x)\varphi(x, f(x))$ is provable in the weaker theory T_2^i , if $i > 0$. This allows us to prove immediately the well-known conservation result for such pairs of theories. As the model-theoretical proof of Theorem 4.27 is rather difficult, we first prove an auxiliary Theorem 4.31 about extensions of models of S_2^{i+1} .

4.19 Definition. Let M be a structure for the language of S_2 , let $A \subseteq M$, $i \geq 0$. Assume that each $\{e\}_i(x)$ defines a function in M . The $\square_i^p$ closure of A is the set

$$\{\{e\}_i(a) \mid a \in A \text{ \& } e \in N\}.$$

We say that A is $\square_i^p$ closed if A is its own closure.

We want to use the fact that *polynomial time predicates are absolute in every $\square_1^p$ closed substructure of M* , a model of S_2^1 . As we do not have symbols for all polynomial time predicates in our language, we have to formalize this statement. Before the next definition let us note that every formula can be transformed to an equivalent formula which has only $\&$, $\vee$ and $\neg$ as connectives, and all negations occur only at atomic formulae. What is important is that this transformation preserves the quantifier complexity. We shall say that such a formula is in *negation normal form*.

4.20 Definition. Let $i \geq 0$, let T be S_2^1 if $i = 0$ and T_2^i otherwise. For a formula φ in negation normal form and a model $M \models T$, we define inductively that φ has $\square_{i+1}^p$ Skolem functions in M :

- (1) every open formula has $\square_{i+1}^p$ Skolem functions in M ;
- (2) if φ and ψ have $\square_{i+1}^p$ Skolem functions in M , then $\forall x\varphi$, $\varphi \& \psi$ and $\varphi \vee \psi$ have $\square_{i+1}^p$ Skolem functions in M ;
- (3) if $\varphi(\mathbf{x}, y)$ has $\square_{i+1}^p$ Skolem functions in M , then $(\exists y)\varphi(\mathbf{x}, y)$ has $\square_{i+1}^p$ Skolem functions in M , if for some f in $\square_{i+1}^p$

$$M \models (\forall \mathbf{x})((\exists y)\varphi(\mathbf{x}, y) \rightarrow \varphi(\mathbf{x}, f(\mathbf{x}))).$$

Here $f(\mathbf{x})$ is represented by $\{\bar{e}\}_i((x_1, (x_2, \dots, x_n) \dots))$, where $e \in N$ and $(x_1, \dots, x_n) = \mathbf{x}$.

4.21 Lemma. Let T be as above, let $M \models T$, let $K \subseteq M$ be $\square_{i+1}^p$ closed, and let $\varphi(\mathbf{x})$ have $\square_{i+1}^p$ Skolem functions in M . Then for $\mathbf{a} \in K$,

$$M \models \varphi(\mathbf{a}) \Rightarrow K \models \varphi(\mathbf{a}).$$

Proof. By induction on the complexity of φ . □

4.22 Corollary. If $\varphi(\mathbf{x})$ and $\neg\varphi(\mathbf{x})$ have $\square_{i+1}^p$ Skolem functions in $M \models T$, T as above, then $\varphi(\mathbf{x})$ is absolute for $\square_{i+1}^p$ closed substructures of M , i.e. if A is a $\square_{i+1}^p$ closed substructure of M and $\mathbf{a}$ is a string of elements of A then

$$M \models \varphi(\mathbf{a}) \quad \text{iff} \quad A \models \varphi(\mathbf{a}).$$

The natural definitions of polynomial time computable predicates have $\square_1^p$ Skolem functions in M . In particular we shall need that the formula $x = (y)_z$ and its negation have $\square_1^p$ Skolem functions in M . Now let $P(x)$ be an arbitrary polynomial time computable predicate. Then it can be defined by

$$(d.1) \quad (\exists w)(|w| \leq p(|x|) \ \& \ \text{Comp}(w, x) \ \& \ (w)_{p(|x|)} = 1).$$

It is clear that w and the existentially quantified numbers in Comp can be computed in polynomial time, thus (d.1) has $\square_1^p$ Skolem functions in M . The same is true for the negation of (d.1), hence (d.1) is absolute. Thus in our formalization, polynomial time predicates are absolute in the situation considered above. The same argument works, of course, for polynomial time computable functions. This enables us to work with polynomial time computable predicates as if they were present in the language.

4.23 Lemma. Let M be a model of S_2^1 , let K be a $\square_1^p$ closed substructure. Then

$$K \models \text{LIND-strict} \hat{\Sigma}_i^b \Rightarrow K \models S_2^1.$$

Proof. First we shall show using *LIND* for *strict* $\hat{\Sigma}_i^b$ formulae that in K every Σ_i^b formula is equivalent to a *strict* $\hat{\Sigma}_i^b$ one. Thus we will have *LIND* Σ_i^b in K . This is proved by induction on the number of quantifiers. Several quantifiers of the same kind are replaced by a single one using the coding relation. A sharply bounded universal quantifier is exchanged with the next existential bounded quantifier using an instance of *BB* $\hat{\Sigma}_i^b$, for which *LIND-strict* $\hat{\Sigma}_i^b$ is sufficient (see the proof of Theorem 4.10). We must check that the properties of the coding function that we are using are provable already in *LIND-strict* $\hat{\Sigma}_i^b$. We need the following properties:

- (1) $(\forall x_0 \dots x_j)(\exists y)((y)_0 = x_0 \ \& \dots \ \& (y)_j = x_j)$, for every j ;
 (2) $(\exists s)(lh(s) = 0)$;
 (3) $(\forall x, s)(\exists t)(lh(t) = lh(s) + \bar{1} \ \& \ (t)_{lh(s)} = x$
 $\ \& \ (\forall z < lh(s))((t)_z = (s)_z))$;

$((t)_{lh(s)})$ is the last element of the sequence t). Again it is clear that these formulae have $\square_1^p$ Skolem functions in M . Since they are true in M they are true also in K by Lemma 4.22.

Similarly we shall show that $LIND \Sigma_i^b$ implies $PIND \Sigma_i^b$ in K . We repeat the proof of

$$S_2^1 + LIND \Sigma_i^b \vdash PIND \Sigma_i^b,$$

(Proposition 4.5(1)) where we needed from S_2^1 only the following properties of the function msp :

$$\begin{aligned} msp(a, \bar{0}) &= \bar{0}; \\ S(b) \leq |a| \rightarrow \perp msp(a, S(b))/2 \perp &= msp(a, b); \\ msp(a, |a|) &= a. \end{aligned}$$

These formulae are true in M and have $\square_1^p$ Skolem functions, hence they are true in K too. $\square$

4.24 Lemma. Let $i \geq 0$, let T be S_2^1 if $i = 0$ and T_2^i otherwise. Then each $\Sigma_0^b(\Sigma_i^b)$ formula has $\square_{i+1}^p$ Skolem functions in each $M \models T$.

Proof. We shall use induction on the depth of the formula. For open formulae the lemma is true by definition. Consider a Σ_i^b formula of the form

$$(\exists x \leq t(a))\varphi(a, x),$$

where we assume that $\varphi(a, x)$ has $\square_{i+1}^p$ Skolem functions. We apply *binary search* using Σ_i^b oracle

$$(\exists x)(p \leq x \leq q \ \& \ \varphi(a, x))$$

to find an x satisfying $x \leq t(a) \ \& \ \varphi(x, a)$, for a given a . At the beginning we set $p := 0$ and $q := t(a)$. First we ask whether $(\exists x \leq t(a))\varphi(a, x)$. If it is not true, then the output is, say, 0. Otherwise we continue as follows. In each subsequent step we set either

$$p := p, \quad q := \perp 1/2(p + q) \perp, \quad \text{if } (\exists x)(p \leq x \leq \perp 1/2(p + q) \perp \ \& \ \varphi(a, x))$$

or

$$p := \lfloor 1/2(p+q) \rfloor, \quad q := q, \quad \text{otherwise.}$$

By Theorem 4.16 the computation of length $|t(\mathbf{a})|$ always exists. In order to prove its correctness we need only to show that in the j -th step of the computation the following two conditions are satisfied:

$$\begin{aligned} q - p &\leq t(\mathbf{a})/2^j; \\ (\exists x)(p \leq x \leq q \ \& \ \varphi(\mathbf{a}, x)). \end{aligned}$$

This is proved by induction on j . Once the computation is given, the condition that we are proving is Σ_i^b , hence we need only $LIND \Sigma_i^b$. If the formula has the form

$$(\exists x \leq |t(\mathbf{a})|) \varphi(\mathbf{a}, x),$$

where φ is Π_i^b , we proceed similarly, but we use thorough search instead of binary search. The case of the universal bounded quantifier and connectives $\&, \vee$ is trivial. $\square$

4.25 Corollary. Let T be as above, let $M \models T$, and let K be a $\square_{i+1}^p$ closed substructure of M . Then K is $\Sigma_0(\Sigma_i^b)$ elementary.

Proof. By Lemma 4.21 and Lemma 4.24. $\square$

Since $LIND \Sigma_i^b$ is a $\Sigma_0^b(\Sigma_i^b)$ formula, we get that K is a model of S_2^i . But we can prove more.

4.26 Theorem. Let $i \geq 0$, let T be S_2^1 if $i = 0$ and T_2^i otherwise. Let K be a $\square_{i+1}^p$ closed substructure of some model M of T . Then K is a model of T_2^i .

Proof. Let $i \geq 0$, let $M \models T$, let $K \subseteq M$ be $\square_{i+1}^p$ closed. We shall show that for every $\alpha(x, \mathbf{b})$ in Σ_i^b , a formula equivalent to $I(\alpha(x, \mathbf{b}))$ (induction for $\alpha(x, \mathbf{b})$) has $\square_{i+1}^p$ Skolem functions, though this is not (known to be equivalent to) a $\Sigma_0^b(\Sigma_i^b)$ formula. Then, by Lemma 4.22, $I(\alpha(x, \mathbf{b}))$ must hold in K . We shall consider the following equivalent formula

$$\neg \alpha(\bar{0}, \mathbf{b}) \vee (\exists x < a)(\alpha(x, \mathbf{b}) \ \& \ \neg \alpha(x + \bar{1}, \mathbf{b})) \vee \alpha(a, \mathbf{b}).$$

Again we shall use binary search. We start with $p := 0$ and $q := a$, and repeat:

$$\begin{aligned} p &:= \lfloor (p+q)/2 \rfloor, \quad q := q, \quad \text{if } \alpha(\lfloor (p+q)/2 \rfloor, \mathbf{b}), \\ p &:= p, \quad q := \lfloor (p+q)/2 \rfloor, \quad \text{otherwise.} \end{aligned}$$

If $\alpha(\bar{0}, b)$ and $\neg\alpha(a, b)$, then this algorithm finds in polynomially many steps (in the length of numbers a, b) some c such that $c < a$, $\alpha(c, b)$, and $\neg\alpha(c + \bar{1}, b)$. It uses Σ_i^b oracle $\alpha(x)$. The existence of the computation follows from Theorem 4.17. We must show the correctness of the algorithm, which means that the value obtained by the computation witnesses the existential quantifier in the induction instance *provably in* T_2^i . We want to show that the following conditions are satisfied in the j -th step of computation:

- (1) $q - p \leq 2^{-j}$;
- (2) $\alpha(p, b)$;
- (3) $\neg\alpha(q, b)$.

We cannot prove by induction that (2) & (3) holds in each step, since this formula has too large complexity. So we prove (1) and (2) using $LIND \Sigma_i^b$ and then we prove separately (3) using $LIND \Pi_i^b$. $\square$

Note that we are not able to prove the theorem with S_2^i instead of T_2^i , since we are not able to formalize $\Box_{i+1}^p$ functions in S_2^i . As a corollary we get a version of the theorem of Buss for fragments T_2^i . It will be a corollary of a stronger result which we shall prove later, but we have to prove it now, since we shall need it for the proof of the stronger one. Moreover the proof is much easier than the proof of Buss's theorem.

4.27 Theorem. Let $i \geq 0$, let $\varphi(x, y)$ be Π_i^b or $strict \hat{\Sigma}_{i+1}^b$ and suppose

$$T_2^i \vdash (\forall x)(\exists y)\varphi(x, y).$$

Then for some f in $\Box_{i+1}^p$

$$N \models (\forall x)\varphi(x, f(x)).$$

Moreover, if T is S_2^1 if $i = 0$ and T_2^i otherwise, then $T \vdash (\forall x)\varphi(x, f(x))$, i.e. more precisely, for some $e \in N$,

$$T \vdash (\forall x)(\exists y)(\{\bar{e}\}_{i+1}(x) = y \ \& \ \varphi(x, y)).$$

Proof. Since for every e , T proves that $\{\bar{e}\}_{i+1}(x) = y$ defines a function, we can use the functional notation. First suppose that φ is Π_i^b . By way of contradiction, suppose that T does not prove $(\forall x)\varphi(x, \{\bar{e}\}_{i+1}\{x\})$ for any e . Then also, for any k , T does not prove

$$\varphi(x, \{\bar{0}\}_{i+1}(x)) \vee \varphi(x, \{\bar{1}\}_{i+1}(x)) \vee \cdots \vee \varphi(x, \{\bar{k}\}_{i+1}(x)),$$

since otherwise, using $\varphi(x, y)$ as an oracle, we could combine these Turing machines into one producing y for a given x . By compactness, there is a model M for

$$T + \neg\varphi(c, \{\bar{0}\}_{i+1}(c)) + \neg\varphi(c, \{\bar{1}\}_{i+1}(c)) + \cdots,$$

where c is a new constant. Let K be the closure of $\{c\}$ under $\square_{i+1}^p$ functions in M , i.e

$$K = \{a \in M \mid (\exists e \in N)(M \models \{\bar{e}\}_{i+1}(c) = a)\}.$$

Then, for every $a \in K$, $M \models \neg\varphi(c, a)$. By Corollary 4.25, K is Σ_i^b elementary in M , hence $K \models (\forall y)\neg\varphi(c, y)$. By Theorem 4.26, $K \models T_2^i$, thus T_2^i does not prove $(\forall x)(\exists y)\varphi(x, y)$.

Now suppose that $\varphi(x, y)$ is *strict* $\hat{\Sigma}_{i+1}^b$. Then for some $\hat{\Pi}_i^b$ formula $\psi(x, z_0, z_1)$ and some term t , $\varphi(x, y)$ is $(\exists z_1 \leq t)\psi(x, y, z_1)$. Hence

$$(\forall x)(\exists y)\varphi(x, y) \equiv (\forall x)(\exists z)(\psi(x, (z)_{\bar{0}}, (z)_{\bar{1}}) \& (z_{\bar{1}} \leq t)).$$

By the first part of the proof z can be witnessed by some $\{\bar{e}\}_{i+1}(x)$, hence y can be witnessed by $(\{\bar{e}\}_{i+1}(x))_{\bar{0}}$, which is a $\square_{i+1}^p$ function. $\square$

4.28 Corollary. Let $i \geq 1$, let $\Phi(x)$ be *strict* $\hat{\Delta}_{i+1}^b$ in T_2^i . Then T_2^i proves induction for $\Phi(x)$.

Proof. The assumption of the theorem means that $\Phi(x)$ is *strict* $\hat{\Sigma}_{i+1}^b$ and for some $\Psi(x)$ *strict* $\hat{\Sigma}_{i+1}^b$, $T_2^i \vdash \Phi(x) \equiv \neg\Psi(x)$. We can represent Φ and Ψ by

$$\begin{aligned}\Phi(x) &\equiv (\exists y)\varphi(x, y), \\ \Psi(x) &\equiv (\exists y)\psi(x, y),\end{aligned}$$

where φ, ψ are in $\hat{\Pi}_i^b$. Thus we have

$$T_2^i \vdash (\forall x)(\exists y)(\varphi(x, y) \vee \psi(x, y)).$$

By Theorem 4.27 we get some e such that

$$T_2^i \vdash \varphi(x, \{\bar{e}\}_{i+1}(x)) \vee \psi(x, \{\bar{e}\}_{i+1}(x)),$$

hence

$$T_2^i \vdash \Phi(x) \equiv \varphi(x, \{\bar{e}\}_{i+1}(x)).$$

We have not gained any reduction of complexity, but this was not our aim, our aim is to witness the induction formula for $\Phi(x)$ using a $\square_{i+1}^p$ function. Suppose we have $\Phi(\bar{0})$ and $\neg\Phi(a)$, and we want to construct some $x < a$ such that $\Phi(x) \& \neg\Phi(x+1)$. We shall use binary search in a similar way as in

Theorem 4.26. We use $\varphi(x, y)$ as a Π_i^b oracle and $\{\bar{e}\}_{i+1}(x)$ as a subroutine in order to ask questions of the form $\varphi(x, \{\bar{e}\}_{i+1}(x))$. The computation goes as follows:

step 0: $p := 0, \quad q := a, \quad u := \{\bar{e}\}_{i+1}(0), \quad v := \{\bar{e}\}_{i+1}(a);$
step $j+1$: $r := \{\bar{e}\}_{i+1}(\lfloor (p+q)/2 \rfloor);$
 if $\varphi(\lfloor (p+q)/2 \rfloor, r)$ then $p := \lfloor (p+q)/2 \rfloor, \quad u := r$
 else $q := \lfloor (p+q)/2 \rfloor, \quad v := r.$

The correctness is proved by $LIND \hat{\Pi}_i^b$ showing that in the j -th step we have

- (1) $q - p \leq 2^{-j}a;$
 (2) $\varphi(p, u) \& \psi(q, v).$ □

A $\exists \Pi_i^b$ formula consists of a prefix of existential quantifiers followed by a Π_i^b formula.

4.29 Theorem. Let $i \geq 0$, let $\varphi(x, y, z)$ be a $\exists \Pi_i^b$ formula and suppose

$$T_2^i \vdash (\forall x)(\exists y)(\forall z)\varphi(x, y, z).$$

Then for some $f_0, \dots, f_n$ in $\square_{i+1}^p$

$$(d.1) \quad N \models (\forall x, z_0, \dots, z_n)(\varphi(x, f_0(x), z_0) \vee \varphi(x, f_1(x, z_0), z_1) \vee \dots \\ \dots \vee \varphi(x, f_n(x, z_0, \dots, z_{n-1}), z_n)).$$

Moreover, this is provable in S_2^1 , if $i = 0$, and in T_2^i , if $i > 0$.

Proof. Let T be S_2^1 , if $i = 0$, and T_2^i , if $i > 0$, let $\varphi(x, y, z)$ be a $\exists \Pi_i^b$ formula. Suppose T does not prove the formula in (d.1) for any choice of $f_0, \dots, f_n$. Take some enumeration of functions in $\square_{i+1}^p$ such that

- (1) the n -th function f_n depends on $\leq n$ arguments;
 (2) each $f \in \square_{i+1}^p$ occurs in the enumeration infinitely many times.

(For instance, $\{e\}_{i+1}(\langle (x_1, x_2), \dots, x_e \rangle)$, where $(-, -)$ is the pairing function, has this property, assuming natural coding of Turing machines.) By compactness,

$$T + \neg\varphi(c, f_0(c), d_0) + \neg\varphi(c, f_1(c, d_0), d_1) + \dots$$

where $c, d_0, d_1, \dots$ are new constants, is a consistent theory. Let M be a model of this theory and let

$$K = \{f_0(c), f_1(c, d_0), f_2(c, d_0, d_1), \dots\}.$$

Since all projections occur in the enumeration and each function occurs in the enumeration infinitely many times, we have

(3) $c, d_0, d_1, d_2, \dots \in K$,

(4) K is $\square_{i+1}^p$ closed.

By (3), we have

$$\forall a \in K \exists d \in K M \models \neg\varphi(c, a, d).$$

Since $\neg\varphi$ is $\forall\Sigma_i^b$, using Corollary 4.25 we get

$$K \models (\forall y)(\exists z)\neg\varphi(c, y, z).$$

□

Now we would like to strengthen Theorem 4.27 by taking the weaker assumption $S_2^{i+1} \vdash (\forall x)(\exists y)\varphi(x, y)$ instead of $T_2^i \vdash (\forall x)(\exists y)\varphi(x, y)$. The proof above does not work for S_2^{i+1} , since we are not able to show that a $\square_{i+1}^p$ closed substructure is a model of S_2^{i+1} . It is worthwhile to realize the reason, then we shall better understand the forthcoming proofs. Let $(\exists y \leq t)\psi(x, y)$ be a Σ_{i+1}^b formula, where ψ is Π_i^b . In a $\square_{i+1}^p$ closed substructure the quantifier $\exists y$ can be witnessed by different elements of the form $\{\bar{e}\}_{i+1}(a)$. Since e runs over standard numbers, we can obtain a cut (for instance contained in some segment $[\bar{0}, b]$) for which $(\exists y \leq t)\psi(x, y)$ holds, hence the induction fails. Therefore we shall take closures under some functions with *nonstandard indices*.

The basic idea of the proof is the same as above: if $(\exists y)\varphi(c, y)$ is not witnessed by a $\square_{i+1}^p$ function, we construct a submodel in which it does not hold. Thus we must be careful when adding $\{e\}_{i+1}(c)$ for e nonstandard. We shall use overspill. If $(\exists y)\varphi(c, y)$ is not witnessed by any $\{e\}_{i+1}(c)$, for e standard, then this must be also true for all e up to some small nonstandard r_1 . Thus we ensure the failure of $(\exists y)\varphi(c, y)$ by taking the closure only under functions with such small indices. However it is not so easy to ensure the induction.

Now we sketch the idea of the model-theoretic proof of this strengthening. It is essentially the proof of Wilkie with some changes. Let $i \geq 0$, let T be S_1^1 if $i = 0$ and T_2^i otherwise. Suppose that for every $e \in N$, T does not prove $\varphi(x, \{\bar{e}\}_{i+1}(x))$. Take a model $M \models T$ with some $c \in M$, such that $M \models \neg\varphi(c, \{\bar{e}\}_{i+1}(c))$. By overspill this is true also for all $e < r_1$. We shall construct substructures consisting of some elements of the form $\{\bar{e}\}_{i+1}(c)$, for $e < r_1$, thus we shall have $\neg(\forall x)(\exists y)\varphi(x, y)$ in the substructures. So we have only to ensure $LIND \Sigma_{i+1}^b$. We shall do it in ω steps (we use only countable models) adding *LIND* for one formula and one string of parameters at a time. Let us take the formula $\Psi(x) \equiv (\exists y \leq t)\psi(x, y)$ considered above, and let a be already in our substructure K . We want to extend K so that it satisfies the following instance of $LIND \Sigma_{i+1}^b$:

$$(d.2) \quad \neg\Psi(\bar{0}) \vee (\exists j \leq |a|)(\Psi(j) \ \& \ \neg\Psi(j + \bar{1})) \vee \Psi(|a|).$$

We take a suitable $r_2 \leq r_1$, and try to find a witness g_0 for $\exists y$ in $\Psi(0)$ of the form $\{e\}_{i+1}(c)$, $e \leq r_2$. If we succeed, we try to find a witness g_1 for $\Psi(1)$ of the form $\{e\}_{i+1}((c, (0, g_0)))$, $e \leq r_2$, and so on. The bound r_2 is chosen so small that we never construct $\{e\}_{i+1}(c)$ with $e \geq r_1$. Suppose, for example, that for some $j < |a|$, we have found g_j , but there is no witness g_{j+1} of the form $\{e\}_{i+1}((c, (j, g_j)))$ for $\Psi(j+1)$. Then we take the $\square_{i+1}^p$ closure of $(c, (j, g_j))$ and K in M , and we have (d.2). We repeat this extension process for other formulae and other parameters. In order to preserve (d.2) we use a similar argument as we used for $\neg(\forall x)(\exists y)\varphi(x, y)$. In our particular case, $\Psi(j)$ is preserved by Σ_i^b elementary extensions, since it is $\exists\Pi_i^b$, while $\neg\Psi(j+1)$ will be preserved because we shall add only elements of the form $\{e\}_{i+1}(c)$, for $e < r_2$. The following key lemma formalizes one step of the above construction.

4.30 Lemma. Let $i \geq 0$, let T be S_2^1 if $i = 0$ and T_2^i otherwise. Let M be a model of T , K substructure of M , $a, b \in K$, let $\Phi(x, y)$ be *strict* $\hat{\Sigma}_{i+1}^b$ and let $\Psi(y)$ be *strict* $\hat{\Pi}_{i+1}^b$. Suppose that

- (1) K is a $\square_{i+1}^p$ closure of one element of M ;
- (2) K is not cofinal in M ;
- (3) $K \models \Psi(b)$.

Then there exists a substructure K^* such that $K \subseteq K^* \subseteq M$, (1)–(3) holds for K^* and

- (4) $K^* \models \neg\Phi(\bar{0}, b) \vee (\exists j < |a|)(\Phi(j, b) \& \neg\Phi(j+1, b)) \vee \Phi(|a|, b)$.

Proof. Let the assumptions be satisfied. Let K be generated by an element c . Since K is not cofinal in M , there is some $d \in M$ such that

$$x \in K \Rightarrow |x| < |d|.$$

Since every $x \in K$ is computed in polynomial time from c , and by overspill, we have

$$(d.3) \quad |c|^{r_0} \leq |d|,$$

for some r_0 nonstandard. Recall that

$$\{e\}_{i+1}(x) = y \equiv \nu_{i+1}(e, x, y, 2^{|x|^\epsilon}),$$

where ν_{i+1} is Δ_{i+1}^b . By (d.3), we can write $\{e\}_{i+1}(x) = y$ as a Δ_{i+1}^b formula with an additional parameter d , since

$$\begin{aligned} M \models \nu_{i+1}(e, x, y, 2^{|x|^\epsilon}) &\equiv (\forall z \leq d)(z = 2^{|x|^\epsilon} \rightarrow \nu_{i+1}(e, x, y, z)) \\ &\equiv (\exists z \leq d)(z = 2^{|x|^\epsilon} \& \nu_{i+1}(e, x, y, z)). \end{aligned}$$

Let

$$\begin{aligned}\Phi(x, \mathbf{b}) &\equiv (\exists z \leq t(x))\varphi(x, z), \\ \Psi(\mathbf{b}) &\equiv (\forall z \leq s)\psi(z),\end{aligned}$$

where $\varphi(x, z)$ is Π_i^b and $\psi(z)$ is Σ_i^b ; we shall omit the parameters $\mathbf{b}$ from now on. Consider the following formula

$$(d.4) \quad (\forall e < |r|)(\{e\}_{i+1}(c) \leq s \rightarrow \psi(\{e\}_{i+1}(x))).$$

Since $K \models \Psi$, K is Σ_i^b elementary in M , and every element of K is of the form $\{e\}_{i+1}(c)$, for e standard, this formula is true in M , for every r standard. Also this formula is equivalent to a Δ_{i+1}^b formula in M . Since $S_2^1 \vdash I\Delta_1^b$, (Theorem 4.7) and for $i \geq 1$, $T_2^i \vdash I\Delta_{i+1}^b$ (Corollary 4.28), we can use overspill to deduce that (d.4) holds in M for some r nonstandard. Let

$$(d.5) \quad r_1 = \min(|r_0|, ||r||, |c|).$$

Now we construct a Turing machine with an oracle which searches for the witness of the existential quantifier in condition (4) of the lemma. The machine works on input c as follows:

- 1: $j := -1$; $g := 0$;
- 2: *find the first $e \leq r_1$ such that*

$$\{e\}_{i+1}((c, (j, g))) \leq t(j + \bar{1}) \ \& \ \varphi(j + \bar{1}, \{e\}_{i+1}((c, (j, g))));$$

- 3: *if such an e does not exist or $j = |a| - 1$, then print (j, g) ;*
- 4: *otherwise $j := j + \bar{1}$; $g := \{e\}_{i+1}((c, (j, g)))$; go to 2 where we define $(-1, 0) = 0$.*

Let e_0 be the code of this Turing machine. We shall estimate e_0 . Recall that it is also the exponent in the time bound of this Turing machine and this makes this estimate nontrivial. To run some $\{e\}_{i+1}$ with $e \leq r_1$ on input c we need at most $|c|^e \leq |c|^{r_1}$ steps. In 2: we do it r_1 times. The time needed to check the Π_i^b condition of 2: is a standard polynomial in $|c|$. (Note that we have to evaluate also the implicit parameters $\mathbf{b}$, for which we need time a standard polynomial in $|c|$.) Finally the cycle given by 4: is repeated $|a| + 1$ times. Thus the total running time is bounded by

$$|c|^{r_1} * r_1 * p(|c|) * (|a| + 1),$$

where $p(x)$ is some standard polynomial. We have $r_1 \leq |c|$ and $|a| + 1$ is also bounded by some standard polynomial in $|c|$, since it is an element of K . Thus we can bound the above expression by $|c|^{2r_1}$. Since the program of the Turing machine has standard length, we can bound its code by

$$e_0 \leq 2r_1.$$

Suppose the machine prints (j, g) on its output. We define K^* as follows. If $K \models \neg\Phi(\bar{0}) \vee \Phi(a)$, then $K = K^*$, otherwise K^* is the $\square_{i+1}^p$ closure of

$$c^* = (c, (j, g)) = (c, \{e_0\}_{i+1}(c)).$$

Now we have to check conditions (2)–(4). We shall assume that $K \models \Phi(\bar{0}) \& \neg\Phi(a)$, hence $M \models \Phi(\bar{0})$, as K is Σ_i^b elementary in M .

We have

$$|c^*| \leq |c|^{2r_1},$$

hence the length of each element x of K^* is bounded by

$$|x| \leq |c|^{k \cdot 2r_1}, \quad k \in N.$$

Since $r_1 \leq |r_0|$, the last expression is bounded by

$$|c|^{r_0} \leq |d|,$$

thus K^* cannot be cofinal in M .

To prove (3), suppose that it fails for K^* , which means that for some $e_1 \in N$ and $h = \{e_1\}_{i+1}(c^*)$,

$$K^* \models h \leq s \& \neg\psi(h).$$

By Corollary 4.25, K^* is Σ_i^b elementary in M , thus this formula holds also in M . Now c^* has been constructed from c using a machine with code $e_0 \leq 2r_1$. Since e_1 is standard, the Turing machine which runs first e_0 and then e_1 has a code e_2 polynomial in r_1 , hence, by (d.5), $e_2 \leq |r|$. But this is in contradiction with (d.4).

It remains to prove condition (4) for K^* . We shall show that

$$K^* \models j < |a| \& \Phi(j) \& \neg\Phi(j + \bar{1}),$$

where (j, g) is the output of the machine e_0 . The first inequality is clear, since the machine stops before reaching $|a|$. Also $\Phi(j)$ is easy, because, by the program lines 2: and 4:,

$$M \models g \leq t(j) \& \varphi(j, g),$$

hence also in K^* by elementariness. By way of contradiction suppose that $K \models \Phi(j + \bar{1})$. This means (again using elementariness) that, for some $e \in N$,

$$M \models \{e\}_{i+1}(c^*) \leq t(j + \bar{1}) \& \varphi(j + \bar{1}, \{e\}_{i+1}(c^*)).$$

In particular $e \leq r_1$. Recall that $c^* = (c, (j, g))$, thus the machine should not stop at j , which is a contradiction. $\square$

4.31 Theorem. Let $i \geq 0$, let T be S_2^1 if $i = 0$ and T_2^i otherwise. Let M be a countable model of T , K a substructure of M , $\mathbf{b} \in K$, and let $\Psi(\mathbf{y})$ be *strict* $\hat{\Pi}_{i+1}^b$. Suppose that

- (1) K is a $\square_{i+1}^p$ closure of one element of M ;
- (2) K is not cofinal in M ;
- (3) $K \models \Psi(\mathbf{b})$.

Then there exists a substructure K^* such that $K \subseteq K^* \subseteq M$ and

- (4) $K^* \models \Psi(\mathbf{b})$;
- (5) $K^* \models S_2^{i+1}$.

Proof. We shall construct a countable chain of substructures

$$K = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq M,$$

and *strict* $\hat{\Pi}_{i+1}^b$ formulae

$$\Psi = \Psi_0, \Psi_1, \Psi_2, \dots$$

Take some enumeration of *strict* $\hat{\Sigma}_{i+1}^b$ formulae with parameters from M . In the j -th step, $j > 0$, we ensure *LIND* in K_j for the first formula $\Phi(a, \mathbf{b})$ from this enumeration for which *LIND* fails in K_{j-1} , and whose parameters $\mathbf{b}$ are in K_{j-1} . Suppose we consider

$$\neg\Phi(\bar{0}, \mathbf{b}) \vee (\exists x < |a|)(\Phi(x, \mathbf{b}) \& \neg\Phi(x + \bar{1}, \mathbf{b})) \vee \Phi(|a|, \mathbf{b})$$

in the j -th step. We apply Lemma 4.30 to K_j , Φ and Ψ_j . Thus we obtain K_{j+1} . Then we define Ψ_{j+1} by

$$\begin{aligned} \Psi_{j+1} &\equiv \Psi_j \& \neg\Phi(\bar{0}), & \text{ if } K_{j+1} \models \neg\Phi(\bar{0}); \\ \Psi_{j+1} &\equiv \Psi_j \& \neg\Phi(k + \bar{1}), & \\ &\text{ if } K_{j+1} \models k < |a| \& \Phi(k) \& \neg\Phi(k + 1), & \text{ for some } k \in K_{j+1}; \\ \Psi_{j+1} &\equiv \Psi_j, & \text{ otherwise.} \end{aligned}$$

Here Ψ_{j+1} is not written in the *strict* $\hat{\Pi}_{i+1}^b$ form; however to show that they are equivalent to such formulae, we need only properties of the coding relation that are expressible by equalities and hence preserved from M to $\square_1^p$ closed substructures. Since Ψ_{j+1} will be true in all K_m , for $m \geq j + 1$, also *LIND*($\Phi(a, \mathbf{b})$) will hold in these models. Consider for instance the second case, where $K_{j+1} \models k < |a| \& \Phi(k) \& \neg\Phi(k + 1)$. Then $\Phi(k)$ will be preserved by extensions, since all the structures are $\square_{i+1}^p$ closed, hence Σ_i^b elementary, while $\neg\Phi(k + \bar{1})$ will be preserved by the definition of the extensions. Let K^* be the union of this Σ_i^b elementary chain. Then all formulae Ψ_j are true in K^* , hence it satisfies condition (4) and

$$K^* \models \text{LIND-strict } \hat{\Sigma}_{i+1}^b.$$

By Lemma 4.23 this implies that K^* is a model of S_2^{i+1} . □

Now we are ready to prove Buss's witnessing theorem. The original theorem was proved with $T = S_2^{i+1}$; the next theorem is a strengthening (since T_2^i is a subtheory of S_2^{i+1}) which is also due to Buss.

4.32 Theorem. Let $i \geq 0$, let $\varphi(x, y)$ be $\hat{\Sigma}_{i+1}^b$. Suppose that

$$S_2^{i+1} \vdash (\forall x)(\exists y)\varphi(x, y).$$

Then for some f in $\square_{i+1}^p$,

$$T \vdash (\forall x)\varphi(x, f(x)),$$

where T is S_2^1 if $i = 0$ and T_2^i otherwise.

Proof. First observe that we can bound y by some term t :

$$S_2^{i+1} \vdash (\forall x)(\exists y \leq t(x))\varphi(x, y).$$

This follows from Theorem 1.4. By Corollary 4.12 we can assume that φ is *strict* $\hat{\Sigma}_{i+1}^b$. Suppose that the conclusion is false. By compactness, there exists a countable model M of T such that

$$M \models \neg\varphi(c, \{e\}_{i+1}(c))$$

for every e standard and moreover, $2^{|c|^k}$, $k \in \mathbb{N}$, is not cofinal in M . Let K be the $\square_{i+1}^p$ closure of the element c in M . Then K is not cofinal in M and

$$K \models (\forall y)\neg\varphi(c, y),$$

since, by Corollary 4.25, it is Σ_i^b elementary in M . Let Ψ be $(\forall y \leq t)\neg\varphi(c, y)$, then it is equivalent to a *strict* $\hat{\Pi}_{i+1}^b$ formula already in predicate logic. Thus we can apply Theorem 4.31 to get some

$$K^* \models S_2^{i+1} + (\forall y \leq t)\neg\varphi(c, y),$$

which is a contradiction. □

The most interesting case is with $i = 0$; we state it explicitly.

4.33 Corollary. Let $\varphi(x, y)$ be a Σ_1^b formula. Suppose that

$$S_2^1 \vdash (\forall x)(\exists y)\varphi(x, y).$$

Then for some f polynomial time computable function

$$S_2^1 \vdash (\forall x)\varphi(x, f(x)).$$

□

4.34 Corollary. For $i \geq 1$, S_2^{i+1} is a $\hat{\Sigma}_{i+1}^b$ conservative extension of T_2^i , i.e. for $\varphi(x)$ in $\hat{\Sigma}_{i+1}^b$,

$$S_2^{i+1} \vdash (\forall x)\varphi(x) \Rightarrow T_2^i \vdash (\forall x)\varphi(x).$$

Proof. Add a dummy existential quantifier to $\varphi(x)$ and apply Theorem 4.32. $\square$

Recall that Δ_i^p sets are sets computable in polynomial time using oracles from Σ_{i-1}^p . Let us note that it is an open problem whether $\Sigma_{i+1}^p \cap \Pi_{i+1}^p = \Delta_i^p$. Thus Δ_i^b definable sets need not be in Δ_i^p .

4.35 Corollary. Let $i \geq 1$, let $\varphi(x)$ be a Δ_i^b formula with respect to S_2^i . Then $\varphi(x)$ defines a Δ_i^p subset of N .

Proof. For $\varphi(x)$ we can find some formulae $\alpha(x, y), \beta(x, y)$ in Π_{i-1}^b and terms t, s such that

$$S_2^i \vdash \varphi(x) \equiv (\exists y \leq t)\alpha(x, y) \equiv \neg(\exists y \leq s)\beta(x, y).$$

Hence

$$S_2^i \vdash (\exists y \leq t)\alpha(x, y) \vee (\exists y \leq s)\beta(x, y).$$

The last formula can be easily transformed into a $\hat{\Sigma}_i^b$ formula; thus we can apply Theorem 4.32 to obtain some f in $\square_i^p$ such that

$$N \models (\forall x)((f(x) \leq t \ \& \ \alpha(x, f(x))) \vee (f(x) \leq s \ \& \ \beta(x, f(x)))).$$

Hence we can decide $\varphi(x)$ by computing $f(x)$ and using a Π_{i-1}^p oracle. $\square$

Again the most interesting particular case of the last corollary is the one with $i = 0$; it can be stated as follows

$$S_2^1 \vdash X \in NP \cap coNP \Rightarrow S_2^1 \vdash X \in P.$$

(e) On the Finite Axiomatizability of Bounded Arithmetic

In this subsection we shall consider the question of the finite axiomatizability of Bounded Arithmetic. First we shall show that the fragments S_2^i and T_2^i are finitely axiomatizable for $i \geq 1$. Thus the finite axiomatizability of S_2 is equivalent to the statement that the hierarchy of these theories is infinite.

Then we shall reduce the finite axiomatizability of S_2 to the non-collapsedness of Polynomial Hierarchy.

4.36 Theorem. For $i \geq 1$, each of S_2^i and T_2^i is finitely axiomatizable.

Proof-sketch. The basic idea of the proof is the same as for fragments IS_n , see Theorem I.2.52. However, since this theorem is important, we give at least a sketch of the proof.

In Theorem 4.18 we showed that there are Σ_i^b formulae μ_i which are in a sense universal; more precisely, for every $\varphi(x)$ in Σ_i^b there exists an e such that

$$S_2^1 \vdash \varphi(x) \equiv \mu_i(\bar{e}, x, 2^{|x|^{\bar{e}}}).$$

Furthermore, $2^{|x|^{\bar{e}}}$ provably exists in S_2^1 ; this follows from the bound

$$2^{|x|^{v+1}} \leq 2^{|x|^v} \# x.$$

For induction and *PIND* we need formulae which have an induction variable and parameters. One parameter is sufficient since the pairing function and the decoding functions are Δ_1^b definable. For the same reason, we can extend the result above to formulae with two free variables:

$$S_2^1 \vdash \varphi(x, y) \equiv \mu_i(\bar{e}, (x, y), 2^{|x|^{\bar{e}}}).$$

Another easy modification of the result above reads as follows:

$$S_2^1 \vdash z \geq 2^{|(x,y)|^{\bar{e}}} \rightarrow (\varphi(x, y) \equiv \mu_i(\bar{e}, (x, y), z)).$$

Now we only need to show that there exists a finite subtheory T of S_2^1 such that for all $\varphi(x, y)$ in Σ_i^b the above equivalence is provable in T . Once we have such a T , we can finitely axiomatize S_2^i by T plus *BASIC* plus

$$(\forall e, y, z) PIND(z \geq 2^{|(x,y)|^e} \rightarrow \mu_i(e, (x, y), z)),$$

where x is the induction variable (thus bounded in *PIND*), and e, y, z are parameters of induction. Similarly we can axiomatize T_2^i .

To find such a T we have to analyze formulae μ_i and ν_i of Theorem 4.18 more closely. The concept that we want to use is just Tarski's conditions for the definition of the satisfaction of Σ_i^b formulae. Tarski's conditions describe the satisfaction of formulae of a given class via relating the truth of a formula with the truth of its components. For each closure condition of the class of formulae there is one Tarski condition. The classes Σ_i^b were defined in Definition 4.2; furthermore we have to use the inductive conditions for the value of terms.

Now we shall describe the conditions more explicitly.

(1) For each function symbol of L_2 we shall have an integer constant, e_S , e_+ , e_* , etc., such that

$$\begin{aligned} S_2^1 &\vdash \nu_0(\bar{e}_S, x, y, 2^{|x|^{\bar{e}_S}}) \equiv y = S(x); \\ S_2^1 &\vdash \nu_0(\bar{e}_+, (x_1, x_2), y, 2^{|(x_1, x_2)|^{\bar{e}_+}}) \equiv y = x_1 + x_2; \end{aligned}$$

etc.

(2) For relation symbols $=$ and $\leq$ we construct binary functions $e_ =$ and $e_ \leq$ such that

$$\begin{aligned} S_2^1 &\vdash \mu_0(e_=(e_1, e_2), x, 2^{|x|^{e_=(e_1, e_2)}}) \\ &\equiv (\exists y_1, y_2)(\nu_0(e_1, x, y_1, 2^{|x|^{e_1}}) \& \nu_0(e_2, x, y_2, 2^{|x|^{e_2}}) \& y_1 = y_2); \end{aligned}$$

and similarly for $\leq$. The functions are Δ_1^b .

(3) For each of the conditions of the definition of the class Σ_i^b we have a function and a condition. Consider for instance the condition

$$\alpha \in \Sigma_i^b \Rightarrow (\forall z \leq |t|) \alpha \in \Sigma_i^b.$$

Then we have a binary Δ_1^b definable function $e_{|V|}$ such that

$$\begin{aligned} S_2^1 &\vdash \mu_i(e_{|V|}(e_1, e_2), x, 2^{|x|^{e_{|V|}}}) \\ &\equiv (\forall y, z)(\nu_0(e_1, x, y, 2^{|x|^{e_1}}) \& z \leq |y| \rightarrow \mu_i(e_2, (x, z), 2^{|(x, z)|^{e_2}})). \end{aligned}$$

The function computed by the machine with the code $e_{|V|}$ does the following. If e_1 is the code of a machine for a term t and e_2 is the code of a machine for a formula $\varphi(x)$, then $e_{|V|}(e_1, e_2)$ is the code of a machine which decides the formula $(\forall z \leq |t|)\varphi(x, z)$.

Let T consist of the above conditions plus finitely many sentences needed to formalize functions $e_ =$, $e_ \leq$, $e_{|V|}$, ... (existence, uniqueness, Δ_1^b definition). Then we can easily show, using induction on the depth of $\varphi \in \Sigma_i^b$, that there exists an e such that

$$T \vdash \varphi(x) \equiv \mu_i(\bar{e}, x, 2^{|x|^{\bar{e}}}).$$

It remains to prove the above Tarski conditions in S_2^1 . A formal proof would, perhaps, require a whole section. But in fact it is not hard to see that they must be provable. Let us look at the last considered condition, which concerns the sharply bounded universal quantifier. Formula μ_i has been defined as being true if the Turing machine with code e accepts input x . The condition describes what the machine with code e of the form $e_{|V|}(e_1, e_2)$ will do: first it will simulate machine e_1 on x and obtain y (the value of the

term), then it will simulate machine e_2 on inputs (x, z) with $z \leq |y|$. Thus we need only to choose the codes of the machines in a suitable way. $\square$

4.37 Corollary. S_2 is finitely axiomatizable if and only if $S_2 = S_2^i$ for some i .

We are not able to decide whether S_2 is finitely axiomatizable. We shall only show that it is not, *assuming a plausible conjecture* that the Polynomial Hierarchy does not collapse on its finite level. This does not seem to be a surprising result because of the relation between the classes of formulae Σ_i^b and complexity classes Σ_i^P . But the relation is not so straightforward. Induction and definability are different things. We shall show that induction is related to optimization and we shall use this fact in our proof.

4.38 Theorem. Let $i \geq 0$ and suppose $\Sigma_{i+2}^P \neq \Pi_{i+2}^P$. Then $T_2^i \neq S_2^{i+1}$.

4.39 Corollary. If the Polynomial Hierarchy does not collapse then

- (i) $S_2^i \neq S_2^{i+1}$ for $i \geq 1$ and $T_2^i \neq T_2^{i+1}$ for $i \geq 0$.
- (ii) S_2 and T_2 are not finitely axiomatizable.
- (iii) $I\Delta_0 + \Omega_1$ is not finitely axiomatizable.
- (iv) $I\Delta_0$ is not finitely axiomatizable.

Proof. (i) and (ii) follow using the inclusions $T_2^0 \subseteq S_2^1 \subseteq T_2^1 \subseteq S_2^2 \dots$.

(iii) $I\Delta_0 + \Omega_1$ is contained in S_2 in such a way that the functions of S_2 are definable in $I\Delta_0 + \Omega_1$.

(iv) Just recall that Ω_1 is a single axiom. $\square$

Before proving Theorem 4.38 formally we shall sketch the main steps of this proof. We shall consider only the case $i = 0$.

Let $C(x, y)$ be a polynomial time predicate. Consider the optimization problem of finding maximal y such that $y \leq x$ and $C(x, y)$. In order to simplify notation we shall assume that the bound $y \leq x$ is implicit in $C(x, y)$ and that $C(x, 0)$ holds for all x . We shall call y a *feasible solution* to x if $C(x, y)$. Let φ be defined by

$$(e.1) \quad \varphi(x, y, z) \equiv C(x, y) \ \& \ (|y| < |z| \rightarrow \neg C(x, z)).$$

Since C is polynomial time computable, φ is Σ_1^b . Now the meaning of $(\forall z)\varphi(x, y, z)$ is that y is the feasible solution to x of maximal size. In S_2^1 the schema $LIND \Sigma_1^b$ proves that for every x there exists a maximal feasible solution y . If $T_2^0 = S_2^1$, then this is true also in T_2^0 , i.e.

$$T_2^0 \vdash (\forall x)(\exists y)(\forall z)\varphi(x, y, z).$$

Now we can apply Theorem 4.29. We obtain some functions $f_0, f_1, \dots, f_n$ computable in polynomial time such that

$$(e.2) \quad N \models (\forall x, z_0, \dots, z_n)(\varphi(x, f_0(x), z_0) \vee \varphi(x, f_1(x, z_0), z_1) \vee \dots \\ \dots \vee \varphi(x, f_n(x, z_0, \dots, z_{n-1}), z_n)).$$

The rest of the proof is only complexity theory. We want to show that such functions cannot exist for all polynomial time predicates C . If n were always 0 (as in Buss's theorem) then our goal would be simple. In such a case, for example, we could construct a maximal clique in a graph using a polynomial time computable function f_0 . This is impossible unless $P = NP$. In fact we can take any NP problem, not only optimization problems. Let $C(x, y)$ be defined, say, by

$$y = 0 \text{ or } y \text{ is a Hamiltonian circuit in graph } x.$$

Then $f_0(x)$ would be 0, if x is not Hamiltonian and $f_0(x)$ would be a Hamiltonian circuit in x otherwise.

But in general n need not be 0. So we modify our problem as follows: x will be a string of graphs and y will be a string of 0's and Hamiltonian circuits in corresponding graphs. Now we assume that for this particular C we obtain (e.2) with $n = 1$, (this is sufficiently instructive). Consider an x , a two element string $x = (G_1, G_2)$, where both graphs are Hamiltonian. Then there are two possibilities:

- (1) $f_0(x)$ is not a feasible solution or $f_0(x) = (y_1, y_2)$ where $y_1 = y_2 = 0$;
- (2) $f_0(x) = (y_1, y_2)$ and y_1 is a Hamiltonian cycle in G_1 or y is a Hamiltonian cycle in G_2 ;

In case (2) f_0 has produced nontrivial information. In case (1) it is not so, but f_1 must produce information: if we take $z_0 = (y', 0)$, where y' is some Hamiltonian cycle in G_1 , then we have $\neg\varphi(x, f_0(x), z_0)$, since either $f_0(x)$ is not a feasible solution or z_0 is a better one. Hence we get

$$(\forall z_1)\varphi(x, f_1(x, z_0), z_1),$$

i.e. $f_1(x, z_0)$ is an optimal solution. This means that $f_1(x, z_0) = (y_1, y_2)$ where both y_1 and y_2 are Hamiltonian cycles. Let us draw an arrow $G_1 \rightarrow G_2$ in case (1) and an arrow $G_2 \rightarrow G_1$ in case (2). This indicates that by having a Hamiltonian circuit for the tail we can construct a Hamiltonian circuit for the head. Consider all Hamiltonian graphs of size m . The above argument shows that there is an arrow at least in one direction between any pair of them. An easy counting argument shows that there exists a set of polynomial size of such graphs which covers the rest. For each of these graphs, we choose a Hamiltonian circuit in it. Let $\mathbb{H}$ be the set consisting of these pairs. $\mathbb{H}$ has also polynomial size. Using $\mathbb{H}$ we can decide in polynomial time whether a graph G is Hamiltonian: try f_0 and f_1 on all pairs (G, H) and (H, G) , where H runs through $\mathbb{H}$. Computations which use additional polynomial size

information are called computations with *polynomial advice*. There is a well-known argument (see [Karp-Lipton 80]) showing that if every *NP* problem can be computed using polynomial time computation with polynomial advice, then $\Sigma_2^P = \Pi_2^P$.

Now we prove Theorem 4.38 in detail. We shall divide the proof into several lemmas. Let $i \geq 0$ be given. A formula $C(x, y)$ will be called a Π_i^b *optimization problem* if $C(x, y)$ is a Π_i^b formula such that $C(x, \bar{0})$ and $C(x, y) \rightarrow y \leq x$ is provable in predicate logic. We shall say that y is a *length optimal* solution to x if the following formula holds:

$$(\forall z)(C(x, y) \& (|y| < |z| \rightarrow \neg C(x, z))).$$

4.40 Proposition. For every Π_i^b optimization problem C , S_2^{i+1} proves that C has a length optimal solution to every x .

Proof. Let $\psi(x, y)$ be defined by

$$\psi(x, u) \equiv (\exists y \leq x)(C(x, y) \& |y| \geq u).$$

By *LIND* Σ_{i+1}^b we have

$$\neg\psi(x, \bar{0}) \vee (\exists u < |x|)(\psi(x, u) \& \neg\psi(x, S(u))) \vee \psi(x, |x|).$$

Since $\psi(x, \bar{0})$ is always true, it just expresses the existence of a length optimal solution. $\square$

It is an easy exercise to prove that, in fact, the existence of length optimal solutions to Π_i^b optimization problems is *equivalent to* *PIND* Σ_{i+1}^b over a sufficiently strong base theory, say S_2^1 .

We are going to use Theorem 4.29. It is convenient to refer to the conclusion of the theorem, see formula (e.2) above, as a kind of interactive way of computing a witness to $\exists y$ in $(\forall x)(\exists y)(\forall z)\varphi(x, y, z)$. Let x be given and suppose we want to construct y . Assume (e.2) holds true. First consider $f_0(x)$. If $(\forall z)\varphi(x, f_0(x), z)$, then take $y = f_0(x)$, otherwise there is some z_0 such that $\neg\varphi(x, f_0(x), z_0)$. We shall call this z_0 a *counterexample* to $f_0(x)$. By (e.2), it must be $n > 0$ and we have

$$N \models (\forall x, z_1, \dots, z_n)(\varphi(x, f_1(x, z_0), z_1) \vee \dots \vee \varphi(x, f_n(x, z_0, \dots, z_{n-1}), z_n)).$$

Hence we can repeat the reasoning above and we get that either $f_1(x, z_0)$ witnesses y or the disjunction can be reduced further by taking a counterexample z_1 to $f_1(x, z_0)$. However this process can be repeated at most n times, since when we obtain

$$(\forall x, z_n)\varphi(x, f_n(x, z_0, \dots, z_{n-1}), z_n)$$

we can stop.

Let $C(x, y)$ be a Π_i^b optimization problem. Assume that $T_2^i = S_2^{i+1}$. Then T_2^i proves that C has always length optimal solutions. This statement can be expressed in the form $(\forall x)(\exists y)(\forall z)\varphi(x, y, z)$ with φ in Σ_{i+1}^b , see (e.1). Hence, by Theorem 4.29, we get that an optimal solution can be computed using $\square_{i+1}^p$ functions and a fixed number of counterexamples. This is our first lemma.

4.41 Lemma. If $T_2^i = S_2^{i+1}$, $i \geq 0$, then for every Π_i^b optimization problem its length optimal solutions can be computed using $\square_{i+1}^p$ functions and a fixed number of counterexamples.

The computations with counterexamples do not define a function complexity class, since we allow the asking of counterexamples only for the particular optimization problem that we want to solve. If we allowed asking arbitrary queries of the same logical complexity, we would obtain just the class $\square_{i+2}^p$. But it is easy to prove that optimal solutions can be computed by such functions. Thus our next step is a reduction to a different function class.

4.42 Definition. A function f belongs to $\square_i^p/poly$ if there exists some g in $\square_i^p$ and a polynomial $p(x)$ such that for every k there is $a_k \leq p(k)$ such that for every x , $|x| = k$,

$$f(x) = g(x, a_k).$$

Here a_k is called a *polynomial advice*; it is some extra information given for free for each size of input; the dependence on the size of input is quite arbitrary.

4.43 Lemma. Let $i \geq 0$. Suppose that for every Π_i^b optimization problem its length optimal solutions can be computed using $\square_{i+1}^p$ functions and a fixed number of counterexamples. Then for every $\psi(x, y)$ in Π_i^p there exists an f in $\square_{i+1}^p/poly$ such that

$$N \models (\forall x)((\exists y \leq x)\psi(x, y) \rightarrow \psi(x, f(x))).$$

Proof. Let ψ be given. Define a Π_i^p optimization problem $\rho(u, v)$ by

$$\rho(u, v) \equiv lh(v) \leq lh(u) \ \& \ (\forall t < lh(v))(\psi((u)_t, (v)_t) \ \& \ (v)_t \leq (u)_t).$$

Let $f_0(u), \dots, f_n(u, v_0, \dots, v_{n-1})$ be some functions in $\square_{i+1}^p$ which interactively compute length optimal solutions for ρ . Let k be given. We shall construct a polynomial advice for inputs of size k . Once we describe the advice, the definition of the function g will be clear. Let

$$V_1 = \{x; |x| = k \ \& \ (\exists y \leq x)\psi(x, y)\}.$$

Choose some function $w(x)$ such that, for $x \in V_1$,

$$\psi(x, w(x)).$$

We shall use the following notation. If $m < lh(u)$ then

$$w(u \mid m)$$

denotes the sequence $(w((u)_0), \dots, w((u)_m))$. To each $n+1$ -tuple u of elements of V_1 , we assign a pair (l, y) , $0 \leq l \leq n$, using the following procedure:

Step 0: compute $f_0(u)$; if $\rho(u, f_0(u))$ and $lh(f_0(u)) > 0$, then put $l = 0$ and $y = (f_0(u))_0$ and stop, otherwise go to step 1;

...

Step m ($m < n$): compute $f_m(u, w(u \mid 0), \dots, w(u \mid m-1))$; if

$$\rho(u, f_m(u, w(u \mid 0), \dots, w(u \mid m-1)))$$

and

$$lh(f_m(u, w(u \mid 0), \dots, w(u \mid m-1))) > m,$$

then put $l = m$ and

$$y = (f_m(u, w(u \mid 0), \dots, w(u \mid m-1)))_m,$$

otherwise go to step $m+1$;

...

Step n: If we have reached this step, then it necessarily holds that

$$\rho(u, f_n(u, w(u \mid 0), \dots, w(u \mid n-1)))$$

and

$$lh(f_n(u, w(u \mid 0), \dots, w(u \mid n-1))) = n+1,$$

thus we put $l = n$ and

$$y = f_n(u, w(u \mid 0), \dots, w(u \mid n-1)),$$

and stop.

Let us call y a *witness* for x if $\psi(x, y)$. The meaning of the above procedure is the following. Let u be an $n+1$ -tuple and let l, y be assigned to it. Then having witnesses for $(u)_0, \dots, (u)_{l-1}$, we can compute the witness y for $(u)_l$.

For an n -element subset Q of V_1 and $x \in V_1 \setminus Q$, we shall say that pair (Q, x) is *good* if for some arrangement $\{x_0, \dots, x_{l-1}, x_{l+1}, \dots, x_n\}$ of Q , l is assigned to u by the above procedure, where u is the sequence

$$(x_0, \dots, x_{l-1}, x, x_{l+1}, \dots, x_n).$$

Define a sequence of subsets of V_1

$$V_1 \supseteq V_2 \supseteq V_3 \cdots,$$

having $N_1, N_2, \dots$, respectively, elements. The $j+1$ -st element in the sequence is defined as follows. Find an n -element subset $Q_j \subseteq V_j$ such that

$$|\{x \in V_j; (Q_j, x) \text{ is good}\}| \geq \frac{N_j - n}{n+1},$$

and take

$$V_{j+1} := V_j \setminus \{x \in V_j; (Q_j, x) \text{ is good}\}.$$

We must show that it is always possible to choose such a Q_j . By the procedure above, for each $n+1$ -element subset $\{x_0, \dots, x_n\}$ of V_j we can construct a good pair (Q, x) such that

$$\{x_0, \dots, x_n\} = Q \cup \{x\},$$

by taking u to be the sequence $(x_0, \dots, x_n)$. Hence there are at least $\binom{N_j}{n+1}$ good pairs. On the other hand there are $\binom{N_j}{n}$ n -element subsets Q of V_j , so at least one such Q must form good pairs with at least

$$\binom{N_j}{n+1} * \binom{N_j}{n}^{-1} = \frac{N_j - n}{n+1}$$

elements. Hence

$$N_{j+1} \leq N_j - \frac{N_j - n}{n+1} = \frac{n}{n+1}(N_j + 1),$$

from which we get

$$\begin{aligned} N_{j+1} &\leq \left(\frac{n}{n+1}\right)^j N_1 + \frac{n}{n+1} + \left(\frac{n}{n+1}\right)^2 + \left(\frac{n}{n+1}\right)^3 + \cdots \\ &< \left(\frac{n}{n+1}\right)^j N_1 + n + 1. \end{aligned}$$

Hence we get $N_t \leq n+2$ after t steps, for

$$t \leq \log_2(N_1) \cdot (\log_2((n+1)/n))^{-1} \leq O(\log_2(2^k)) = O(k).$$

We take the polynomial size advice to be the sequence of pairs $(x, w(x))$, where x runs through all elements of

$$Q_1 \cup Q_2 \cup \cdots \cup Q_t \cup V_t.$$

Now let x be such that

$$|x| = k \ \& \ (\exists y \leq x) \psi(x, y),$$

i.e. $x \in V_1$. Then either $x \in V_t$ and so a witness for x is in the advice, or (Q_j, x) is a good pair for some $j \leq t$. But then we can construct a witness for x from the witnesses for the elements of Q_j (using $\square_{i+1}^p$ functions in the procedure above). $\square$

4.44 Lemma. Let $i \geq 0$. Suppose that for every $\psi(x, y)$ in Π_i^p there exists an f in $\square_{i+1}^p/poly$ such that

$$N \models (\forall x)((\exists y \leq x) \psi(x, y) \rightarrow \psi(x, f(x))).$$

Then $\Sigma_{i+2}^p = \Pi_{i+2}^p$.

Proof. Let $A(b)$ be in Π_{i+2}^p . Without loss of generality we can assume that it can be represented in the form

$$(\forall x)(|x| = |b| \rightarrow (\exists y \leq b) \gamma(b, x, y)),$$

where γ is Π_i^b .

Let $\psi(x, y)$ be a Π_i^b formula such that

$$\psi((b, x), y) \equiv |x| = |b| \rightarrow (y \leq b \ \& \ \gamma(b, x, y)).$$

Let f be a function in $\square_{i+1}^p/poly$ guaranteed by Lemma 4.43 for ψ , and let function g in $\square_{i+1}^p$ and polynomial p be guaranteed by Definition 4.42 for f . Then we can write $A(b)$ in the following Σ_{i+2}^b form¹

$$A(b) \equiv (\exists a)(a \leq p(|(b, b)|) \ \& \ (\forall x)(|x| = |b| \rightarrow \gamma(b, x, g(a, (b, x))))).$$

The right to left implication is trivial, the left to right implication follows by taking an advice a . $\square$

This finishes the proof of Theorem 4.38.

There is a similar reduction of the question $S_2^i = T_2^i?$ to a problem in complexity theory. First we shall define that for an optimization problem $C(x, y)$, y is an *optimal* solution to x if

$$(\forall z)(C(x, y) \ \& \ (y < z \rightarrow \neg C(x, z))).$$

Secondly, we define interactive computations with an unbounded number of counterexamples as the the natural extension of the concept above. Of course,

¹ Note that g is Π_{i+1}^b definable by Theorem 4.17.

there is an implicit polynomial bound to the number of counterexamples, since we consider only polynomial time oracle computations. Note that the number of feasible solutions to x may be exponential in $|x|$, thus polynomial number of counterexamples apparently cannot help to solve some difficult optimization problems (such as the TRAVELLING SALESPERSON), while it gives trivial algorithms if we ask only for *length* optimal solutions (as in CLIQUE for instance).

We have the following counterpart of Theorem 4.29.

4.45 Theorem. Let $i \geq 0$, let $S_2^i \vdash (\forall x)(\exists y)(\forall z \leq t)\varphi(x, y, z)$, for φ in Σ_{i+1}^b and t a term. Then, for a given x we can compute y using $\Box_{i+1}^p$ computations with (an unbounded number of) counterexamples.

This theorem can be proved using the proof theoretical method which Buss used for his witnessing theorem. We are not going to prove it here. The counterpart of Lemma 4.41 is the following.

4.46 Corollary. For $i \geq 1$, if $S_2^i = T_2^i$, then the optimal solutions for every Π_{i-1}^b problem can be computed using $\Box_{i+1}^p$ computations with (an unbounded number of) counterexamples.

We conjecture that the conclusion of Corollary 4.46 is false, hence also that $S_2^i \neq T_2^i$.

*

The research on witnessing functions is still going on and we can look for new nice results. The task of proving that Bounded Arithmetic is not finitely axiomatized *without assumptions on complexity classes* seems still too hard. What might be more accessible is to prove nonconservation results for S_2^i and T_2^i using such assumptions.

5. Interpretability and Consistency

(a) Introduction

In the last section we were mainly interested in the strength of theories obtained by restricting the quantifier complexity of bounded formulae in the scheme of induction. The main theme of this section is investigation of the strength of theories obtained from $I\Sigma_0$ by adding functions of a different growth rate. (These functions have graphs definable by Σ_0 formulae, so the strength is increased by assuming that they are total.) We shall

consider typical questions of metamathematics: interpretations, provability of consistency, conservativity, etc.

First we shall consider the definition of truth for Σ_0 formulae in Bounded Arithmetic. As a consequence we obtain that $I\Sigma_0 + Exp$ is finitely axiomatizable. Another consequence (which uses further technical results) is the existence of an interpretation of $I\Sigma_0 + \Omega_n$ in Q , $n = 1, 2, \dots$, which in turn implies that Q and these theories are equiconsistent in $I\Sigma_0 + \Omega_1$. For further results we need a formalization of a weaker form of the cut-elimination theorem in $I\Sigma_0 + Exp$. The usual form of this theorem is unprovable in this theory. On the other hand Gödel's incompleteness theorem can be strengthened further for such fragments. For instance $I\Sigma_0 + Exp$ does not prove the consistency of Q . This and several other results of this kind will be shown in subsection (f). Finally we shall consider the question of limited use of exponentiation in $I\Sigma_0$.

(b) Truth Definitions for Bounded Formulae

In Chap. I, Sect. 1 we have shown that truth for Σ_0 formulae can be defined by a Δ_1 formula in $I\Sigma_1$. Because of diagonalization there is no such Σ_0 definition. However we can omit the main $\exists$ quantifier in the Σ_1 formula and thus we obtain a Σ_0 formula with an additional parameter. Then we estimate how large the parameter must be; put another way, we are looking for a suitable bound to the $\exists$ quantifier in the Σ_1 definition of Σ_0 truth. Sometimes we shall call the Σ_0 formula obtained in this way a *truth definition* for Σ_0 formulae, though it is not quite precise. The construction will be essentially the same as in Chap. I, only we have to compute the bounds more precisely.

Before constructing a formula with a parameter which defines the truth for Σ_0 formulae, we shall briefly consider this problem from the point of view of complexity theory. Assume that there is a definition of Σ_0 truth described above; suppose that it has the quantifier complexity k and the bound to the parameter is some function f . Then we have the following relation between complexity classes $\Sigma_0^N \subseteq \Sigma_k^{[f]}$ (see Sect. 2 for definitions). Thus the corresponding question is: for which function g and constant k do we have $\Sigma_0^N \subseteq \Sigma_k^g$? Recall that Σ_0^N are just the sets in the Linear Time Hierarchy (Theorem 2.16), hence the best that we can achieve (to our present knowledge) is

$$\Sigma_0^N \subseteq \bigcup_{c \in N} Time(c^n) = E.$$

This result can be used to write a definition of truth for Σ_0 formulae based on a Σ_0 formula with an exponential bound to the *length* of the parameter, but again we not only need a formula, but also proofs of its properties in $I\Sigma_0$. Furthermore we need an estimate of the constant c . Therefore we present an explicit construction.

We shall use the notation introduced in Sect. 3. Thus for instance $\max(s)$ is the maximal element of sequence s . In this subsection we use only the standard arithmetical language L_0 . The formalization of syntax is as described in Sect. 3; thus if x is the Gödel number of an expression, $|x|$ is its length.

Our approach is to consider a bounded formula literally as a program. The computation of terms needs no explanation; the interpretation of bounded quantifiers is as search procedures. The crucial thing is to estimate the size of the numbers that will occur in such a computation.

Let us consider an example first. Let $\alpha(x)$ be a bounded formula

$$(\exists y \leq x^2)(\exists z \leq y^2)(\exists u \leq z^2)\beta(x, y, z, u),$$

where β is also a bounded formula. To compute the truth value of $\alpha(x)$ we use three *FOR* loops, one for each bounded variable. Note that the third loop has the range 0 to x^{2^3} .

5.1 Proposition ($I\Sigma_0$). (1) For a term t and a string of numbers s ,

$$t(s) \leq (\max(s) + 2)^{|t|}.$$

(2) The maximal value needed to compute the truth of a bounded formula α on a string s is bounded from above by

$$(\max(s) + 2)^{2^{|\alpha|}}.$$

Note in passing that, by (1), the value of a *closed* term t is bounded by $2^{|t|} \leq 2t$, hence it can be defined as a total function in $I\Sigma_0$, while for general terms in L_0 we need $I\Sigma_0 + \Omega_1$. Nevertheless the estimates above are provable in $I\Sigma_0$ provided that the upper bound exists.

Proof. (1) is easily proved by induction on the length of t .

(2) This is also proved by induction on the length of α . We shall compute only the induction step for a quantifier. Let $\alpha(s)$ be $(\exists y \leq t(s))\beta(y, s)$ (for $\forall$ it is the same). Assume we have bound

$$(\max(s, y) + 2)^{2^{|\beta|}},$$

for $\beta(y, s)$. By (1) we know that

$$t(s) \leq (\max(s) + 2)^{|t|},$$

hence for $\alpha(s)$ we get

$$\begin{aligned} (\max(s, (\max(s) + 2)^{|t|}) + 2)^{2^{|\beta|}} &= (((\max(s) + 2)^{|t|}) + 2)^{2^{|\beta|}} \\ &\leq ((\max(s) + 2)^{|t|+1})^{2^{|\beta|}}. \end{aligned}$$

We shall estimate the exponent:

$$(|t| + 1)2^{|\beta|} \leq 2^{|t|+1+|\beta|} \leq 2^{|\alpha|},$$

which proves the induction step. $\square$

First we define the graph of the *val* function which defines the value of arithmetical terms. Also here we need an additional parameter. The meaning of *Value*(y, x, z, u) is: y is the value of term x computed on string z , u is the parameter. Let *Var*(x, i) be a formalization of the concept of the i -th variable.

5.2 Proposition. There exists a Σ_0 formula *Value*(y, x, z, u) and a polynomial p such that it is provable in $I\Sigma_0$ that:

- (1) $Term(x_1) \& Term(x_2) \& |x_1|, |x_2| \leq v \& |u| > p(|\max(z)|, v) \rightarrow$
 $\rightarrow (Var(x_1, i) \rightarrow Value(y, x_1, z, u) \equiv (z)_i = y)$
 $\& (Value(y, \bar{0}, z, u) \rightarrow y = \bar{0})$
 $\& (Value(y_1, S^\bullet(x_1), z, u) \& Value(y_2, x_1, z, u) \rightarrow y_1 = S(y_2))$
 $\& (Value(y, x_1 +^\bullet x_2, z, u) \& Value(y_1, x_1, z, u)$
 $\& Value(y_2, x_2, z, u) \rightarrow y = y_1 + y_2)$
 $\& (Value(y, x_1 *^\bullet x_2, z, u) \& Value(y_1, x_1, z, u)$
 $\& Value(y_2, x_2, z, u) \rightarrow y = y_1 * y_2);$
- (2) $Term(x) \& |x| \leq v \& |u_1|, |u_2| \geq p(|\max(z)|, v)$
 $\& Value(y_1, x, z, u_1) \& Value(y_2, x, z, u_2) \rightarrow y_1 = y_2.$

Proof. *Value*(y, x, z, u) is defined as $(\exists s \leq u)\varphi(y, x, z, s)$, where $\varphi(y, x, z, s)$ is a Σ_0 formula expressing the following conditions:

- (1) s is a sequence of pairs $(a_0, b_0), \dots, (a_m, b_m)$;
- (2) for every $i \leq m$, either a_i is a variable, or a_i is $a_j +^\bullet a_k$, where $0 \leq j, k < i$, etc. for $S, \bar{0}$ and $*$;
- (3) if a_i is the j -th variable, then $b_i = (z)_j$; if a_i is $a_j +^\bullet a_k$, then $b_i = b_j + b_k$; etc. for $S, \bar{0}$ and $*$;
- (4) $y = b_m$ and $x = a_m$.

Clearly, if s satisfying these conditions exists, then it determines y uniquely and the conditions of the proposition are provable. (Namely, using induction, one can prove that every sequence of pairs for a subterm of t is a subsequence of some sequence of pairs for t .) Thus we only have to find an upper bound to the size of s . Assume s is given. By Proposition 1 (1),

$$b_i \leq (\max(z) + 2)^{|a_i|}.$$

Since the pairing function is a polynomial, (a_i, b_i) is bounded by a polynomial in v and $(\max(z) + 2)^v$, hence $(a_i, b_i) \leq (\max(z) + 2)^{q(v)}$ for some polynomial q . By Proposition 3.30, $|s| \leq r(|\max(z) + 2|^{q(v)}, v)$ for some polynomial r , which gives the required bound $|s| \leq p(|\max(z)|, v)$, p a polynomial.

The details can be worked out in a similar way as we did in Sect. 3. $\square$

5.3 Corollary. The value of an L_0 term is definable by a Σ_0^{exp} formula (in fact by a bounded formula in L_2).

Proof. Define

$$val(x, z) = y \equiv_{df} Value(y, x, z, 2^{p(|\max(z)|, |x|)}). \quad \square$$

The meaning of $\Gamma(x, z, u)$ in the following theorem is: the Σ_0 formula x is satisfied by a string of numbers z and u is some bound.

5.4 Theorem. There exists a Σ_0 formula $\Gamma(x, z, u)$ and a constant c such that it is provable in IS_0 that:

$$\begin{aligned} |u| \geq (\max(z) + 2)^{c^v} \ \&x \ |x| \leq v. \rightarrow (x = t =^{\bullet} s \rightarrow (\Gamma(x, z, u) \equiv \\ &(\exists y)(Value(y, t, z, u) \ \& \ Value(y, s, z, u)))) \ \& \\ &\&x (x = (x_1 \rightarrow^{\bullet} x_2) \rightarrow (\Gamma(x, z, u) \equiv (\Gamma(x_1, z, u) \rightarrow \Gamma(x_2, z, u)))) \ \& \\ &\&x (x = \neg^{\bullet} x_1 \rightarrow (\Gamma(x, z, u) \equiv \neg \Gamma(x_1, z, u))) \ \& \\ &\&x (x = (\exists^{\bullet} w \leq^{\bullet} t) x_1 \rightarrow (\Gamma(x, z, u) \equiv (\exists y, y')(Value(y', t, z, u) \ \& \ y \leq y' \\ &\& \ \& \ \Gamma(x_1, z', u))))), \end{aligned}$$

where we assume that w is the j -th variable for some j and sequence z' is obtained by replacing its j -th entry by y ; the conditions (Tarski's conditions) for $\leq^{\bullet}$ and bounded $\forall^{\bullet}$ are similar and therefore omitted.

Proof. The proof is similar to the one above. We define $\Gamma(x, z, u)$ as $(\exists s \leq u) \psi(s)$, where $\psi(s)$ says that s is a sequence consisting of pairs of some substitution instances of subformulae and their truth values satisfying Tarski's conditions. Let s be the sequence of pairs $(a_0, b_0), \dots, (a_m, b_m)$. We shall require the following condition. Let a_i be of the form $(\exists w \leq t) \varphi(w)$ or $(\forall w \leq t) \varphi(w)$. Then we include all formulae $\varphi(\bar{n})$ for $n \leq val(t, z)$ in $a_0, \dots, a_m$. Here $\bar{n}$ denotes the dyadic numeral introduced in Sect. 3(g). (If the value of $(\exists^{\bullet} w \leq^{\bullet} t) \varphi(w)$ turns out to be TRUE, then it suffices to include just one $\varphi(\bar{n})$, but this is not a real advantage.) By Proposition 1(2), the values needed to compute the truth value of a formula x are bounded by

$$(\max(z) + 2)^{2^{|x|}}.$$

Thus the a_i 's are Gödel numbers of subformulae of x with some variables replaced by numerals $\bar{n}$ for

$$n \leq (\max(z) + 2)^{2^{|x|}}.$$

The length of such formulae is estimated by

$$c_1 * |(\max(z) + 2)^{2^{|x|}}| * |x|,$$

where c_1 is a constant, since the length of a numeral is linear in its value. The length of (a_i, b_i) is thus

$$c_2 * |(\max(z) + 2)^{2^{|x|}}| * |x|,$$

for some constant c_2 . Now we estimate the number m of such pairs. We have at most $|x|$ variables in x , and each variable can be replaced by a numeral $\bar{n}$ with the bound above for n . This gives $((\max(z) + 2)^{2^{|x|}} + 1)^{|x|}$ possibilities and there are at most $|x|$ subformulae, hence

$$m \leq |x| * ((\max(z) + 2)^{2^{|x|}} + 1)^{|x|}.$$

Thus we can estimate the length of s by

$$c_3 * |(\max(z) + 2)^{2^{|x|}}| * |x|^2 * ((\max(z) + 2)^{2^{|x|}} + 1)^{|x|} \leq (\max(z) + 2)^{c^{|x|}},$$

where c_3 and c are suitable constants. Hence if u is larger than or equal to this bound, then there exists some sequence s that witnesses the truth of x . The provability of Tarski's conditions is shown as above. $\square$

For a string of variables $z_1, \dots, z_n$ we denote by $(z_1, \dots, z_n)$ the code of the sequence $z_1, \dots, z_n$.

5.5 Corollary. (1) For every bounded formula $\alpha(z_1, \dots, z_n)$ in L_0 there exists a constant k such that

$$I\Sigma_0 \vdash u \geq 2^{(\max(z_1, \dots, z_n) + 2)^k} \rightarrow (\alpha(z_1, \dots, z_n) \equiv \Gamma(\bar{\alpha}, (z_1, \dots, z_n), u)).$$

(2) Let M be a model of $I\Sigma_0$, let $a, b, d \in M$, let $\alpha(x)$ be a Σ_0 formula. Suppose d is nonstandard and $b > 2^{(a+2)^d}$. Then

$$M \models \alpha(a) \equiv \Gamma(\bar{\alpha}, (a), b).$$

Proof. (1) First apply Proposition 5.2 to prove the statement for α atomic using induction on the complexity of terms in it. Then for general α use induction on the logical complexity of α and Theorem 5.4.

(2) This is an immediate consequence of (1). $\square$

5.6 Theorem. $I\Sigma_0 + Exp$ is finitely axiomatizable.

Proof. Let T be a finite fragment $I\Sigma_0$ such that T proves Tarski's conditions of Theorem 5.4 for Γ . Then for every α bounded, T proves the formula of Corollary 5.5 (1). Let φ be defined by

$$\varphi(x, y, p, u) \equiv \Gamma(x, (y, p), u).$$

(Here we are misusing notation: to be quite precise, we should distinguish between a *pair* and a *two element sequence*, but we do not want to introduce new notation.) Let T' be T plus the least number principle for $\varphi(x, y, p, u)$, where y is the induction variable and x, p and u are parameters, plus Exp . Let $\alpha(y, p)$ be an arbitrary bounded formula for which we want to prove the least number principle (clearly it suffices to have just one parameter). We shall argue in T' . Let p be given and suppose that for some y_0 we have $\alpha(y_0, p)$. Since we have Exp , we can take

$$u \geq 2^{(\max(y_0, p))^k},$$

where k corresponds to α . Hence, for $y \leq y_0$,

$$\alpha(y, p) \equiv \Gamma(\bar{\alpha}, (y, p), u) \equiv \varphi(\bar{\alpha}, y, p, u).$$

Thus we can take the least y for $\varphi(\bar{\alpha}, y, p, u)$. □

(c) An Interpretation of $I\Sigma_0$ in Q

Q is a very weak theory, e.g. the associativity of $+$ is not provable in it. However from the point of view of interpretability it is quite strong. We shall show that $I\Sigma_0 + \Omega_n$ is interpretable in Q for every n . Moreover the form of interpretation is very natural: we take the same operations and only restrict the domain. The domain of interpretation is, roughly speaking, an initial segment of the original numbers. This suggests an attractive finitist's program, which was pursued by Edward Nelson [Nelson 86]. Namely, the consistency of Q seems quite evident; further it is also evident that interpretation preserves consistency, thus if we develop mathematics only in theories interpretable in Q , we are safe-guarded against inconsistencies. Later on we shall see that we can even interpret in Q some consequences of $I\Sigma_0 + Exp$ which are not available in $I\Sigma_0 + \Omega_n$ (Theorem 5.27 (i)).¹

¹ Let us remark, for those who are interested in minimal foundations of mathematics, that one can use an extremely weak system for set theory instead of Q , since Q is interpretable in it. The system has only two simple axioms:

$$\begin{aligned} &(\exists x)(\forall y)(\neg y \in x); \\ &(\forall x)(\forall y)(\exists z)(\forall u)(u \in z \equiv (u \in x \vee u = y)). \end{aligned}$$

The main reason why we consider interpretation of Bounded Arithmetic in Q is, however, different. The interpretability of systems of Bounded Arithmetic gives us *equiconsistency* of them with Q , and this equiconsistency is provable in Bounded Arithmetic too. Hence Gödel theorem implies that such a system does not already prove consistency of Q .

The rest of the subsection is almost entirely devoted to the proof of the following theorem. In the whole subsection we shall use only interpretations given by restricting the domain. Such an interpretation is determined by a formula with one free variable.

5.7 Theorem. For every n , there exists a global interpretation of $I\Sigma_0 + \Omega_n$ in Q .

By a *global* interpretation we mean the usual concept of interpretation. The word *global* is used here to stress the fact that we have one translation of the language (but in fact only the domain will be different) such that the translations of each one of the infinitely many axioms are provable. On the other hand, *local* interpretation means that we can choose a different translation for every finite subset of axioms. Nelson uses local interpretations in his book. We shall use a local interpretation as an intermediate step.

We shall say that a formula $I(x)$ is *inductive* in a theory T if

$$T \vdash I(\bar{0}) \ \& \ (\forall x)(I(x) \rightarrow I(S(x))).$$

Recall that I is a *cut* in T if it satisfies moreover

$$I(x) \ \& \ y \leq x \rightarrow I(y),$$

provably in T . We shall say that $J(x)$ is a *subcut* of $I(x)$ in T if moreover

$$T \vdash J(x) \rightarrow I(x).$$

Let Q^+ be Q augmented with the following axioms:

$$\begin{array}{ll} (\text{associativity of } +) & (x + y) + z = x + (y + z); \\ (\text{left distributivity}) & x(y + z) = xy + xz; \\ (\text{associativity of } *) & (xy)z = x(yz). \end{array}$$

5.8 Lemma. There exists an inductive formula $I(x)$ which determines an interpretation of Q^+ in Q .

We shall not prove this lemma. It can be proved using similar tricks as we shall use below, but it is technically more complicated, since we start with a very weak theory Q . A complete proof can be found in [Nelson 86].

5.9 Lemma. Let $I(x)$ be an inductive formula in Q . Then there exists a subcut $J(x)$ of $I(x)$ in Q .

Proof. We shall use the following three easy theorems of Q :

- (i) $x \leq \bar{0}$,
- (ii) $x \leq y \equiv S(x) \leq S(y)$,
- (iii) $\bar{0} \leq x$.

Let I be inductive in Q . Let K be defined by

$$K(x) \equiv I(x) \& (\forall y, z)(z \leq y \& y \leq x \rightarrow z \leq x).$$

We shall show that K is inductive. $K(\bar{0})$ follows easily from (i) and from $I(\bar{0})$. Suppose $K(x)$. Then, clearly, $I(S(x))$ and it remains to prove the second part of the condition for $K(S(x))$. Let $z \leq y \leq S(x)$. If $z = \bar{0}$, then, by (iii), $z \leq x$. Otherwise, by (Q3), $z = S(z')$, for some z' . Then, by (i), (Q5) and (Q1), y cannot be $\bar{0}$. Thus $y = S(y')$, for some y' . Thus we get $z' \leq y' \leq x$, using (ii). By $K(x)$, we have $z' \leq x$. Using (ii) once again, we get $z = S(z') \leq S(x)$. Hence K is inductive.

Now define

$$J(x) \equiv (\forall y \leq x)K(y).$$

First we shall show that J is inductive. $J(\bar{0})$ follows easily from (i). Assume $J(x)$ and let $y \leq S(x)$. We need to show that $K(y)$. If $y = \bar{0}$, then $K(y)$, since K is inductive. Suppose that $y = S(y')$, for some y' . Then, by (ii), we get $y' \leq x$, hence $K(y')$. Since K is inductive, $K(S(y'))$, which is $K(y)$. Thus K is inductive. Now we shall prove that

$$J(x) \& y \leq x \rightarrow J(y).$$

Suppose $J(x) \& y \leq x$. Let $z \leq y$ be arbitrary. We need to prove $K(z)$. By $J(x)$, we have in particular $K(x)$, hence $z \leq x$. Thus we have $K(z)$, since $J(x)$. $\square$

5.10 Lemma. Let $I(x)$ be an inductive formula in Q^+ . Then there exists a cut $J(x)$ such that J is a subcut of I closed under S , $+$ and $*$, i.e. Q^+ proves

$$\begin{aligned} J(x) &\rightarrow I(x); \\ J(\bar{0}); \\ J(x) \& J(y) &\rightarrow J(S(x)) \& J(x+y) \& J(xy). \end{aligned}$$

Proof. By Lemma 5.9 we may assume that I is a cut. We shall use the same construction as in the proof of Theorem 3.5, Chap. III. Take

$$\begin{aligned} J_0(x) &\equiv (\forall y)(I(y) \rightarrow I(y+x)); \\ J(x) &\equiv (\forall y)(J_0(y) \rightarrow J_0(yx)). \end{aligned}$$

Then J_0 contains $\bar{1}$, since I is inductive; J_0 is closed under $+$ by associativity of $+$; J is contained in J_0 (since J_0 contains $\bar{1}$) and hence also in I ; J is closed under $+$ by left distributivity and since J_0 is closed under $+$; J is closed under S , since it contains $\bar{1}$ and it is closed under $+$; J is closed under $*$ by associativity of $*$. To show that J is closed downwards, we use the assumption that I is a cut, left distributivity and J_0 being closed under $+$. $\square$

5.11 Lemma. $I\Sigma_0$ is locally interpretable in Q .

Proof. By Lemma 5.8 we can take Q^+ instead of Q . Let $\varphi_1(x, \mathbf{p}), \dots, \varphi_n(x, \mathbf{p})$ be given bounded formulae. We want to interpret Q plus induction for $\varphi_1(x, \mathbf{p}), \dots, \varphi_n(x, \mathbf{p})$ in Q^+ . Let

$$I_j(x, \mathbf{p}) \equiv \varphi_j(\bar{0}) \& (\forall y \leq x)(\varphi_j(y, \mathbf{p}) \rightarrow \varphi_j(S(y), \mathbf{p})) \rightarrow \varphi_j(x, \mathbf{p}),$$

for $j = 1, \dots, n$. Let

$$I(x) \equiv (\forall \mathbf{p})(I_1(x, \mathbf{p}) \& \dots \& I_n(x, \mathbf{p})).$$

Since each $I_j(x, \mathbf{p})$ is inductive with respect to x in Q^+ , so is $I(x)$. By Lemma 5.10 we can find a subcut $J(x)$ closed under $+$ and $*$. Since J is closed under S , $+$ and $*$ and it is closed downwards, we have, for any bounded formula $\alpha(z_1, \dots, z_k)$,

$$Q^+ \vdash J(z_1) \& \dots \& J(z_k) \rightarrow (\alpha(z_1, \dots, z_k) \equiv (\alpha(z_1, \dots, z_k))^J).$$

Since

$$J(x) \rightarrow I(x) \rightarrow I_j(x)$$

provably in Q^+ , for every j , J determines an interpretation of induction for every φ_j . All but two axioms of Q (even Q^+) are open, hence they hold in the interpretation. The remaining two axioms are (Q3): $x \neq \bar{0} \rightarrow (\exists y)(x = S(y))$ and (Q8): the definition of the relation $\leq$. The first one is provable using induction for bounded formulae from the others (hint: first prove $x \leq S(x)$, then $x \neq \bar{0} \rightarrow (\exists y \leq x)(x = S(y))$). Hence we can simply suppose that we have enough formulae amongst $\varphi_1, \dots, \varphi_n$ to prove it. The last one causes no problems, since we can trivially interpret Q in the theory obtained from Q by eliminating $\leq$ from the language and deleting (Q8). $\square$

Proof of Theorem 5.7. By Theorem 3.5, Chap. III, we know that each cut in a sufficiently strong theory can be shortened to a cut closed under ω_n , for a given n . One can easily check that $I\Sigma_0$ is strong enough for that. So, by the last lemma, we only have to interpret $I\Sigma_0$ in a finite fragment T of $I\Sigma_0$. We take T so strong that

- (1) the properties of the exponentiation relation are provable in it,
- (2) Tarski's conditions for the formula Γ of Theorem 5.4 are provable in it, and
- (3) the least number principle for $\Gamma(x, (y, p), u)$, with y as the induction parameter, is provable in it.

By the same theorem of Chap. III on shortening cuts we get a cut $I(x)$ such that

$$T \vdash I(x) \rightarrow (\exists y)(y = 2^{2^x}).$$

By Lemma 5.10, we can assume that I is closed under $+$ and $*$. We claim that I determines an interpretation of $I\Sigma_0$. It suffices to prove the least number principle for every bounded $\alpha(y, p)$ with just one parameter. Let p be an element in I and suppose that for some y_0 in I we have $\alpha(y_0, p)$. Take

$$u = 2^{(\max(y_0, p))^k},$$

where k corresponds to α . This is possible, since

$$2^{(\max(y_0, p))^k} \leq 2^{2^{\max(y_0, p)^{k*k}}},$$

and the last number exists by a property of I . Now we conclude as in Theorem 5.6: for $y \leq y_0$,

$$\alpha(y, p) \equiv \Gamma(\bar{\alpha}, (y, p), u),$$

and we can take the least y for $\Gamma(\bar{\alpha}, (y, p), u)$. □

The concept of consistency *Con* used in the following theorem is the usual one as used in Chap. III, except that we must use the efficient coding of sequences as described in Sect. 3 of this section.

5.12 Theorem. For every n ,

$$I\Sigma_0 + \Omega_1 \vdash \text{Con}(I\Sigma_0 + \Omega_n^*) \equiv \text{Con}(Q^*).$$

Proof. The consistency of $I\Sigma_0 + \Omega_n$ implies the consistency of Q trivially. The consistency of $I\Sigma_0 + \Omega_n$ can be reduced to the consistency of Q using an interpretation. We shall analyze this proof, in order to see that it can be done in $I\Sigma_0 + \Omega_1$. Let a proof of a contradiction in $I\Sigma_0 + \Omega_n$ be given. Using the interpretation constructed above we can transform it into a proof of a contradiction in Q . Again, it is clear that the interpretation as a relation can be described by a Σ_0 formula. Hence we only have to check that the size of the transformed proof is polynomial in the original one. This can be proved by showing that each application of a logical rule in the original proof is replaced by only a polynomially longer sequence in the new proof. There

is only one place where we have to modify the interpretation given above. When we prove the least number principle for a bounded formula α , we define parameter u by

$$u = 2^{(\max(y_0, p))^k}.$$

Now we have to take into account how k depends on α , since α is not fixed. Thus instead of the double exponential function in the definition of $I(x)$, we take triple exponential. Hence

$$Q \vdash I(x) \rightarrow (\exists y) (J(y) \& y = 2^{2^{2^x}}),$$

where cut $J(y)$ determines an interpretation of a sufficiently strong finite fragment T of $I\Sigma_0$ in Q . Then we replace number k in the definition of u by a more explicit bound given by Theorem 5.4:

$$u = 2^{(\max(y_0, p))^{c|\alpha|}},$$

where c is the constant independent of α from Theorem 5.4. Now, using the estimate

$$2^{(\max(y_0, p))^{c|\alpha|}} \leq 2^{2^{\max(y_0, p) + c|\alpha|}}$$

one can show that u exists and $J(u)$ holds true, and the proofs of these facts in Q are of polynomial length in the length of α . Hence the induction axioms are replaced also by polynomially long sequences. $\square$

(d) Cut-Elimination and Herbrand's Theorem in Bounded Arithmetic

There are two basic measures of the complexity of proofs: the length of proofs and the quantifier complexity of proofs. Cut-elimination and Herbrand's theorem implies that each first order tautology α has a proof whose quantifier complexity is not larger than the quantifier complexity of α . The cost for the reduction of the quantifier complexity is an increase in the size of the proof. This is the reason why this theorem is not provable even in $I\Sigma_0 + Exp$ in full generality. Nevertheless there is a restricted version of this result which is provable in $I\Sigma_0 + Exp$ and which has several applications. We shall prove it in this subsection. The proof will be just an analysis of a classical proof of cut-elimination. There is an alternative proof based on Hilbert style calculi, but we shall use a variant of Gentzen's approach since it is simpler and more transparent. The transformations used in the proof will again be some simple manipulations with sequences, hence definable by Σ_0 formulae. Thus we only have to estimate the size of constructed proofs.

We shall use a slightly simplified version of Schwichtenberg's system [Schwichtenberg 77] and follow his proof of cut-elimination. The proofs will

be only briefly sketched; however we do not presuppose knowledge of proof theory. For more details the reader should consult [Takeuti 80]. It should be stressed here that the translations from this system to the system introduced in Chap. 0 and backwards are quite elementary. In particular (assuming efficient coding, which was introduced in Sect. 3), it is possible to prove in $IS_0 + \Omega_1$ that a sentence is provable in one system iff its translation is provable in the other one.

*

We consider a system for first order logic. Its language consists of logical connectives $\&$, $\vee$, $\neg$, quantifiers $\forall$, $\exists$, variables, constants, function symbols and relation symbols. We do not distinguish equality as a special (or logical) relation symbol. In order to reduce the number of rules, and consequently the number of cases to be considered in the proofs, we allow negation only at atomic formulae. Thus, for a complex formula φ , the expression $\neg\varphi$ is an abbreviation for the equivalent formula obtained by applying De Morgan rules. For a similar reason we shall treat disjunction $\vee$ differently too. We shall think of the formulae in a disjunction as forming a set rather than a sequence. In the arithmetization, all syntactical objects are represented as sequences (which in turn are represented as binary expansions). Thus what we mean is that: *in proofs we allow the replacement of a disjunction by another one which has the same elements*. For reasons of symmetry we shall consider conjunction $\&$ to be also a set operation. When we write, for instance, that a formula φ has the form $\gamma \vee \psi$, we mean that φ is the disjunction of a set $\{\gamma_1, \dots, \gamma_m, \psi_1, \dots, \psi_n\}$, where $\gamma_1, \dots, \gamma_m, \psi_1, \dots, \psi_n$ are atomic formulae, conjunctions or formulae starting with a quantifier, and we do not assume that the sets $\{\gamma_1, \dots, \gamma_m\}$, $\{\psi_1, \dots, \psi_n\}$ are disjoint and γ is the disjunction of $\gamma_1, \dots, \gamma_m$, or $\gamma = \gamma_1$ if $m = 1$, and ψ is the disjunction of $\psi_1, \dots, \psi_n$, or $\psi = \psi_1$ if $n = 1$.

A *proof* is a sequence of formulae (usually called *proof lines*, or *steps*) where each formula is either a logical axiom or follows from preceding ones by a rule.

There is one axiom schema in our system:

$$(A) \quad \gamma \vee \varphi \vee \neg\varphi,$$

and five rules:

$$(W) \quad \frac{\gamma}{\gamma \vee \varphi}, \quad (\&) \quad \frac{\gamma \vee \varphi, \gamma \vee \psi}{\gamma \vee (\varphi \& \psi)},$$

$$(\forall) \quad \frac{\gamma \vee \varphi(x)}{\gamma \vee (\forall y)\varphi(y)} \text{ where the variable } x \text{ is not free in } \gamma,$$

$$(\exists) \quad \frac{\gamma \vee \varphi(\tau)}{\gamma \vee (\exists x)\varphi(x)} \text{ where } \tau \text{ is a term free for } x \text{ in } \varphi,$$

$$(\text{Cut}) \frac{\gamma \vee \varphi, \delta \vee \neg \varphi}{\gamma \vee \delta}.$$

The elements of the disjunctions γ and δ are called *side formulae*; γ and δ may consist of one formula or can even be void. In *Cut*, φ and $\neg \varphi$ are *cut formulae*. Let us point out that an important rule, the *contraction rule*, is only implicit in our system. A typical application of it, which is also needed in the proof of Herbrand's theorem, is the following: we derive by $(\exists)$

$$\frac{\gamma \vee (\exists x)\varphi(x) \vee \varphi(\tau)}{\gamma \vee (\exists x)\varphi(x)}.$$

In the lower sequent we do not repeat $(\exists x)\varphi(x)$ once more, since our disjunctions are *sets*. Rule (W) is clearly redundant (and in contrast with (Cut) also harmless), but it is convenient to have it in the system.

The fact that we define proofs to be *sequences*, not *trees* (the latter being common in proof theory), will be important later, when we shall consider systems weaker than $I\Sigma_0 + \text{Exp}$.

We shall follow the notation of Chap. IV, Sect. 4 by writing $\text{Proof}(d, \gamma)$ for the relation d is a proof of γ . By writing $\gamma(x)$ and then $\gamma(\tau)$, we denote the substitution of τ for all free occurrences of x in γ . We shall use two measures of the complexity of proofs. The *length*, denoted $lh(d)$, is the length of the sequence which codes the proof d . Of course, lh is defined also for formulae and terms. We assume that the number of symbols used in the coding sequences is finite, thus the Gödel number of a proof d is bounded by $c^{lh(d)}$, for some constant c . Further we define, for a formula φ , $rank(\varphi)$ as the number of connectives and quantifiers in φ plus 1. The *cut-rank* of a proof d , denoted by $c\text{-rank}(d)$, is 0 if there is no cut in it; otherwise it is the maximum of the ranks of the cut formulae.

5.13 Lemma ($I\Sigma_0 + \text{Exp}$). (i) Suppose $\text{Proof}(d, \gamma \vee (\forall y)\psi(y))$; then there exists d' such that $\text{Proof}(d', \gamma \vee \psi(x))$ and $c\text{-rank}(d') \leq c\text{-rank}(d)$, $lh(d') \leq lh(d)$.

(ii) If $\text{Proof}(d(x), \gamma(x))$, then for any term τ , $\text{Proof}(d(\tau), \gamma(\tau))$ and $c\text{-rank}(d(x)) = c\text{-rank}(d(\tau))$, $lh(d(\tau)) \leq lh(d(x)) * lh(\tau)$.

(iii) Suppose $\text{Proof}(d, \varphi_0 \& \varphi_1)$, then there exist d_0, d_1 such that $\text{Proof}(d_i, \varphi_i)$ and $c\text{-rank}(d_i) \leq c\text{-rank}(d)$, $lh(d_i) \leq lh(d)$, for $i = 0, 1$.

(We take x and τ in (i) and (ii) so that there are no clashes of variables.)

Intuitively, the statements are clear; formally they can be proved by induction on the length of the proof. The proofs are easy, therefore we omit them.

5.14 Reduction Lemma ($I\Sigma_0 + \text{Exp}$). Suppose d is a proof of Θ such that the last inference is a cut with a cut formula φ . Suppose that the part of the proof before the last inference has cut-rank less than $rank(\varphi)$. Then there exists d' , a proof of Θ , such that $c\text{-rank}(d') < c\text{-rank}(d)$ and $lh(d') \leq lh(d)^3$.

Proof. Let the last inference be

$$\frac{\gamma \vee \varphi, \delta \vee \neg \varphi}{\gamma \vee \delta}.$$

1. First suppose that φ is atomic. Then $\text{rank}(\varphi) = 1$, hence there is no other cut in the proof. The proof d contains a proof d_0 of $\gamma \vee \varphi$ of rank 0. Consider how φ enters the proof d_0 . If it is by (W) or as a side formula in (A), then we can omit it and derive $\gamma \vee \delta$ by (W). If it is by (A) of the form $\sigma \vee \varphi \vee \neg \varphi$, then we first derive $\sigma \vee \delta \vee \neg \varphi$ from $\delta \vee \neg \varphi$ by weakening (W) and then use this proof instead of the proof line $\sigma \vee \varphi \vee \neg \varphi$. We remove φ also from the proof lines below $\sigma \vee \varphi \vee \neg \varphi$ and extend them by forming disjunctions with δ . Thus we transform the proof of $\gamma \vee \varphi$ into a proof d' of $\gamma \vee \delta$.

2. Suppose φ is $(\exists x)\psi(x)$. Our strategy is the same as above, but now φ can be introduced also by rule ($\exists$). (Note that φ can be introduced in d several times by ($\exists$); see the example above describing the contraction.) Consider such a derivation in d :

$$\frac{\sigma \vee \psi(\tau)}{\sigma \vee (\exists x)\psi(x)},$$

where it is possible that the lower disjunction is just σ . In d we have a proof of $\delta \vee \neg \varphi$, which is $\delta \vee (\forall x)(\neg \psi(x))$. By Lemma 5.13 (i) and (ii) we get a proof of $\delta \vee \neg \psi(\tau)$. Now we apply cut to $\sigma \vee \psi(\tau)$ and $\delta \vee \neg \psi(\tau)$ to get $\sigma \vee \delta$. This cut has smaller rank, hence, after all these changes, we get a proof d' of $\gamma \vee \delta$ of a smaller rank.

3. If φ is a conjunction $\varphi_0 \& \varphi_1$, then we do the same procedure as in 2 except that we now use Lemma 5.13 (iii).

4. By duality there are no other cases. If, for instance, φ is a disjunction, then $\neg \varphi$ is a conjunction.

5. It remains to compute the size of the transformed proof d' . This will also be a proof that the above considerations can be carried out in $I\Sigma_0 + \text{Exp}$. The proof d' is constructed in such a way that for certain proof lines of d we add a proof of $\delta \vee \neg \varphi$ or a proof of $\delta \vee \neg \psi(\tau)$, or a proof of $\delta \vee \varphi_i$ or just extend the disjunction by δ and remove φ from it. The largest increase in the size occurs when we need the proof of $\delta \vee \neg \psi(\tau)$; there we have the following upper bound from Lemma 5.13:

$$lh(d) * lh(\tau) \leq lh(d')^2.$$

As there are at most $lh(d)$ places where we do such extensions, the bound is $lh(d)^3$. $\square$

5.15 Lemma ($I\Sigma_0 + \text{Exp}$). If $\text{Proof}(d, \gamma)$ and $c\text{-rank}(d) > 0$, then there exists d' such that $\text{Proof}(d', \gamma)$, $c\text{-rank}(d') < c\text{-rank}(d)$ and $lh(d') \leq 2^{2^{*lh(d)}}$.

Proof. Take the initial segment of proof d down to the first cut whose rank is equal to $c\text{-rank}(d)$. Replace this subproof by a proof whose rank is less than $c\text{-rank}(d)$ using Lemma 5.14. Repeat this procedure until no cut with rank equal to $c\text{-rank}(d)$ remains. We start with a proof of length $lh(d)$; then in each step the length is raised to the 3, and there are at most $lh(d)$ cuts that we have to eliminate. This gives

$$lh(d') \leq lh(d)^{3^{lh(d)}} \leq 2^{2^{2 \cdot lh(d)}}. \quad \square$$

Now we shall define several concepts of provability and consistency. The basic concepts have already been defined in Chap. III; we repeat their definition for the reader's convenience.

5.16 Definition.

(0) d is a proof of x in T

$$\begin{aligned} \text{Proof}_T(d, x) \equiv_{df} (\exists s \leq d)((\forall i < lh(s))T((s)_i) \& \\ \& \text{Proof}(d, (s)_0 \&^\bullet \dots \&^\bullet(s)_{lh(s)-1} \rightarrow^\bullet x)). \end{aligned}$$

(1) *Provability in the theory T*

$$\text{Pr}_T(x) \equiv_{df} (\exists d)\text{Proof}_T(d, x).$$

(2) *Restricted provability in the theory T*

$$\text{RPr}_T(k, x) \equiv_{df} (\exists d)(c\text{-rank}(d) \leq k \& \text{Proof}_T(d, x)).$$

(3) *Cut-free provability in the theory T*

$$\text{CFPr}_T(x) \equiv_{df} \text{RPr}_T(\bar{0}, x).$$

(4) *Herbrand provability in the theory T*

$$\begin{aligned} \text{HPr}_T(x) \equiv_{df} \text{“there exists a propositional proof of} \\ \text{a disjunction of instances of the open part of} \\ \text{He}(\bigwedge S \rightarrow x), \text{ for some finite } S \subseteq T'’, \end{aligned}$$

(see 0.18 for the definition of the function He).

(5) *Consistency of the theory T*

$$\text{Con}_T \equiv_{df} \neg \text{Pr}_T(\overline{0 = 1}).$$

(6) *Restricted consistency of the theory T*

$$\text{RCon}_T(k) \equiv_{df} \neg \text{RPr}_T(k, \overline{0 = 1}).$$

(7) *Cut-free consistency of the theory T*

$$\text{CFCon}_T \equiv_{df} \neg \text{CFPr}_T(\overline{0 = 1}).$$

(8) *Herbrand consistency of the theory T*

$$\text{HCon}_T \equiv_{df} \neg \text{HPr}_T(\overline{0 = 1}).$$

(If we omit the subscript, it means *provability* or *consistency* in pure logic.)

The theories that we shall use will be either finite or have naturally defined polynomial time computable sets of axioms. For such theories T the predicate “ φ is an axiom of T ” can be expressed by a Σ_1^b formula. Then $\text{Proof}_T(d, x)$ is also a Σ_1^b formula.

Denote by *Superexp* an axiom saying that the Σ_0 definable function 2_x^x is total (see Definition III.3.3).

5.17 Theorem

- (i) $I\Sigma_0 + \text{Superexp} \vdash (\forall x)(\text{Pr}^\bullet(x) \rightarrow \text{CFPr}^\bullet(x)).$
- (ii) For every k , $I\Sigma_0 + \text{Exp} \vdash (\forall x)(\text{RPr}^\bullet(\bar{k}, x) \rightarrow \text{CFPr}^\bullet(x)).$

Proof. Both statements follow immediately from Lemma 5.15. □

Let us note that it is provable in $I\Sigma_0 + \text{Exp}$ that each *propositional* tautology has a cut-free proof. This can be shown by formalizing a completeness proof for the propositional fragment of our system consisting of the schema (A) and rule (&). Let a propositional tautology φ be given. Construct a proof of φ by applying rule (&) backwards, i.e. we split conjunctions into parts until only disjunctions of atomic and negated atomic formulae are left (recall that negations can occur only at atomic formulae). We must show that these disjunctions are of the form (A) which just means that they are tautologies. This is true because rule (&) is sound also in the opposite direction. In each step of this construction we add two formulae of length at most $lh(\varphi)$ and the construction has at most $lh(\varphi)$ steps. Thus the total size of the proof is estimated from above by $lh(\varphi) * 2^{lh(\varphi)}$. Therefore, the construction can be carried out in $I\Sigma_0 + \text{Exp}$.

* *

We conclude this subsection by showing that in $I\Sigma_0 + Exp$ the cut-free provability and the Herbrand provability are equivalent. Then we derive corollaries about the provability of the consistency of Q .

5.18 Middle Sequent Theorem. Let φ be a first order tautology in *prenex normal form*. Then there exists a proof d of φ which is cut-free and such that it can be divided into two parts d' and d'' such that d' is quantifier free, while d'' contains only applications of the quantifier rules.

The idea of the **proof** is simple. First one shows that axiom schema (A) and rule (W) can be restricted so that the newly introduced formulae must be quantifier-free. This increases the length of proofs only polynomially and no new cuts are needed. Now take such a cut-free proof of φ and switch quantifier and non-quantifier rules to obtain the required form of a proof. The only thing that we have to take care of are possible collisions of variables in rule ($\forall$). Therefore we must first transform the proof into a *regular proof*, which is a proof in which each application of rule ($\forall$) has its own variable that is generalized (*eigenvariable*). To do such a renaming we have to transform the proof into tree form, which means that each premise is used only once. This may cause at most an exponential increase. Hence it is provable in $I\Sigma_0 + Exp$ that a formula φ has such a proof whenever it has a cut-free proof.

Suppose such a proof d of φ is given. Let d' be the quantifier-free initial segment of the proof, let φ' be its last formula. We may also suppose that each formula, except the last one, is a premise of some application of a rule in d . Hence φ is obtained from φ' by successive applications of quantifier rules. In this situation we call φ' the *middle formula* of d . (Usually in proof theory sequents are used instead of our use of disjunction of formulae; therefore the theorem is referred to as the Middle Sequent Theorem.) We shall show that φ' can be easily transformed into a tautological Herbrand disjunction $He(\varphi)$.

Let φ be of the form

$$(\exists x_1)(\forall y_1) \dots (\exists x_n)(\forall y_n) \psi(x_1, y_1, \dots, x_n, y_n).$$

We can transform any formula in a prenex normal form into such a form by adding dummy quantifiers. Then φ' has the following form

$$\bigvee_i \psi(t_1^i, z_1^i, \dots, t_n^i, z_n^i),$$

where t_j^i are some terms and z_j^i are variables. Let $f_1, \dots, f_n$ be new function symbols that will correspond to the universally quantified variables of φ . Now we follow the derivation of φ from φ' , but instead of applying the rules, we substitute terms in φ' as follows. If generalization ($\forall$) is applied to the variable z_j^i , then we substitute $f_j(t_1^i, \dots, t_j^i)$ for z_j^i . If the rule ($\exists$) is used, then we

do nothing. In order to prove that the resulting disjunction is a Herbrand disjunction, we shall show two things:

- (i) $t_1^i, \dots, t_j^i$ are changed (i.e. something is substituted in them) only after the generalization of z_j^i ;
- (ii) if $z_j^i = z_{j'}^k$, then $j = j'$ and $t_1^i = t_1^k, \dots, t_j^i = t_j^k$.

Suppose (i) is false. Then some z_m^k which is generalized before z_j^i must occur in $t_1^i, \dots, t_j^i$. Clearly z_m^k is not among $z_1^i, \dots, z_j^i$. Hence before z_m^k can be generalized, its occurrences in $t_1^i, \dots, t_j^i$ must be killed by existentially quantifying the terms in which it occurs. But this is a contradiction, since such existential quantifications are done only after the universal quantification of z_j^i .

Let us prove (ii). If z_j^i occurs in different disjuncts, they must be first contracted into one before we can generalize (see the restriction on $(\forall)$). Two different disjuncts may become identical in the course of the derivation d'' after we replace some terms and free variables in them by bounded variables. However, different $t_1^i, \dots, t_j^i$ and $t_1^k, \dots, t_j^k$ can be identified in such a way only after we quantify z_j^i (by the same argument as in (i)). Thus (ii) must be true.

Now it is clear that the procedure changes φ' gradually into a Herbrand disjunction. Condition (i) ensures that the terms have the required form, condition (ii) ensures that we replace the same variables by the same terms, hence we preserve the property of being a tautology. In each step the size of the disjunction increases quadratically (since we substitute some substrings of a string into itself), hence the size of the resulting $He(\varphi)$ can be estimated by

$$lh(d)^{2^{lh(d)}} \leq 2^{2^{2^{lh(d)}}}.$$

The converse transformation, from $He(\varphi)$ into a cut-free proof of φ , is also easy. Consider a formula φ of the same form as above. First eliminate the new function symbols $f_1, \dots, f_n$ in $He(\varphi)$ by replacing maximal terms starting with such a function symbol by new variables. Let us denote the resulting formula by φ' . Since $He(\varphi)$ is a *propositional* tautology, so is φ' . Take some proof of φ' ; we have shown above that there is always one that is cut-free and at most exponentially long. Now we only have to see that we can apply the quantifier rules in a suitable order to suitable terms and variables to get φ . The following rules lead to a proof of φ :

- (1) Whenever we have a disjunct of the form

$$(\forall y_j)(\exists x_{j+1})(\forall y_{j+1}) \dots (\exists x_n)(\forall y_n) \\ \psi(t_1, z_1, \dots, t_j, y_j, x_{j+1}, y_{j+1}, \dots, x_n, y_n),$$

apply $(\exists)$ to the term t_j ;

(2) if the above is not possible, then consider formulae of the form

$$(\exists x_{j+1})(\forall y_{j+1}) \dots (\exists x_n)(\forall y_n) \\ \psi(t_1, z_1, \dots, t_j, z_j, x_{j+1}, y_{j+1}, \dots, x_n, y_n);$$

take one of them such that the term in $He(\varphi)$ to which z_j corresponds is of maximal length, and generalize z_j .

If we repeat this rule the disjunctions clearly converge to φ . Thus we only have to show that the rule can be applied repeatedly, until the disjunction becomes φ . For $(\exists)$ there are no restrictions, for $(\forall)$ we have to show that z_j occurs in only one disjunct and it does not occur in $t_1, z_1, \dots, t_{j-1}, z_{j-1}, t_j$ of this disjunct. The last thing is clear, since $z_1, \dots, z_j$ correspond to the terms $f_1(t_1), \dots, f_j(t_1, \dots, t_j)$, hence $f_j(t_1, \dots, t_j)$ cannot be a subterm of any of the terms

$$t_1, f_1(t_1), \dots, t_{j-1}, f_{j-1}(t_1, \dots, t_{j-1}), t_j.$$

Now suppose (1) is not possible and suppose z_j occurs in another disjunct. Since (1) is not possible, the disjunct must have the following form

$$(\exists x_{k+1})(\forall y_{k+1}) \dots (\exists x_n)(\forall y_n) \\ \psi(t'_1, z'_1, \dots, t'_k, z'_k, x_{k+1}, y_{k+1}, \dots, x_n, y_n).$$

The term corresponding to z'_k is $f_k(t'_1, \dots, t'_k)$. If z_j occurred in $t'_1, z'_1, \dots, t'_k$, then the term corresponding to z'_k would be longer than the one corresponding to z_j . Hence we can have only $z_j = z'_k$. But then $f_j(t_1, \dots, t_j) = f_k(t'_1, \dots, t'_k)$, thus $j = k$, and $t_1 = t'_1, \dots, t_j = t'_k$. Hence it is the same formula.

We have proved the following theorem.

5.19 Theorem. It is provable in $I\Sigma_0 + Exp$ that: a formula has a cut-free proof if and only if it has a Herbrand proof. In symbols

$$I\Sigma_0 + Exp \vdash (\forall x)(CFPr^\bullet(x) \equiv HPr^\bullet(x)).$$

By formalizing a classical proof of the consistency of arithmetic we get the following result.

5.20 Theorem.

- (i) $I\Sigma_0 + Superexp \vdash Con_{I\Sigma_0}$.
- (ii) $I\Sigma_0 + Exp \vdash HCon^\bullet_Q$ and for every k , $I\Sigma_0 + Exp \vdash RCon^\bullet_Q(\bar{k})$.

Proof. First we transform Q into an open theory Q' : we delete axiom (Q8) defining $\leq$, add the predecessor function and replace axiom (Q3) by an open

axiom. Assume $\neg HCon(Q')$, hence we have some propositional tautology $He(\neg Q')$, which is a disjunction of substitution instances of $\neg Q'$. But this is impossible, since in $I\Sigma_0 + Exp$ we can define the truth for open formulae. Thus we have $I\Sigma_0 + Exp \vdash HCon_{Q'}^*$.

To prove (i), first reduce the consistency of $I\Sigma_0$ to the consistency of Q' using the interpretations of $I\Sigma_0$ in Q (Theorem 5.12) and Q in Q' . Then apply cut-elimination in $I\Sigma_0 + Superexp$ (Theorem 5.17(i)), and the equivalence of the cut-free and Herbrand consistency (Theorem 5.19).

$RCon_{Q'}^*(\bar{k})$ is reduced to $RCon_{Q'}^*(\bar{m})$, for some m , by the interpretation of Q in Q' . Applying cut elimination (Theorem 5.17(ii)), this is further reduced to $CFCOn_{Q'}^*$, and by Theorem 5.19 to $HCon_{Q'}^*$. Thus we get

$$I\Sigma_0 + Exp \vdash RCon_{Q'}^*(\bar{k}).$$

Now $RCon_{Q'}^*(\bar{0})$ is just $CFCOn_{Q'}^*$ which is equivalent to $HCon_{Q'}^*$ in $I\Sigma_0 + Exp$. This shows the rest of (ii). $\square$

(e) The Π_1 Theorems of $I\Sigma_0 + Exp$

In this subsection we prove a theorem of A. Wilkie which characterizes Π_1 sentences provable in $I\Sigma_0 + Exp$ as the Π_1 sentences interpretable in Q and consider some related problems. We start with some lemmas. Recall that x denotes the formalized numeral function and that we are using dyadic numerals of Sect. 3.

5.21 Lemma. For every inductive formula $I(x)$ in Q , there exists k such that

$$S_2^1 \vdash (\forall x) RPr_{Q'}^*(\bar{k}, I(x)).$$

Proof. Recall that a numeral $\bar{n}$, for $n > 0$, is constructed as follows. We take $n_0, \dots, n_k$ such that $n_0 = 1, n_k = n$ and $n_{i-1} = \lfloor n_i/2 \rfloor$. Then $\bar{n}_0$ is $\bar{1}$ and, for $i = 1, \dots, k$, $\bar{n}_i$ is

$$\begin{aligned} \bar{2} * \bar{n}_{i-1}, & \quad \text{if } n_i = 2n_{i-1}; \\ (\bar{2} * \bar{n}_{i-1}) + \bar{1}, & \quad \text{if } n_i = 2n_{i-1} + 1; \end{aligned}$$

where $\bar{1}$ is $S(\bar{0})$ and $\bar{2}$ is $(S(\bar{0}) + S(\bar{0}))$. For a given inductive $I(x)$, take some $J(x)$ which is a subcut of I and $J(x) \rightarrow J(\bar{2} * x)$, provably in Q . Then to show $J(\bar{n})$ we need at most $2k + 1$ proof steps. The formulae that we use are of polynomial size in the length of n , hence the whole proof has polynomial size. Therefore S_2^1 can formalize this proof. Note that the rank of formulae used in the proof does not depend on $\bar{n}$, hence is determined only by I . $\square$

5.22 Lemma. There exists k such that for every two terms $t(x_1, \dots, x_n)$ and $s(x_1, \dots, x_n)$ in L_0 :

$$\begin{aligned} S_2^1 \vdash (\forall x_1, \dots, x_n) ((t(x_1, \dots, x_n) = s(x_1, \dots, x_n) \rightarrow \\ RPr_Q^\bullet(\bar{k}, t(\dot{x}_1, \dots, \dot{x}_n) = s(\dot{x}_1, \dots, \dot{x}_n))) \\ \& (t(x_1, \dots, x_n) \neq s(x_1, \dots, x_n) \rightarrow \\ RPr_Q^\bullet(\bar{k}, t(\dot{x}_1, \dots, \dot{x}_n) \neq s(\dot{x}_1, \dots, \dot{x}_n))))). \end{aligned}$$

In the following proof and the next lemma we shall denote the function assigning the dyadic numeral to a number also by *numeral*(x); the dot notation would be rather confusing in some formulae.

Proof. Let $I(x)$ be a cut in Q on which the operations $+$ and $*$ have the properties of ring operations. We can use this cut to construct proofs of instances of associativity, commutativity, etc. for numerals. Namely, by the lemma above, we first show that the numerals in question belong to I and then we apply the law in I . This cut will be the only factor that influences the cut-rank k of the resulting proof. Now the sentence

$$RPr_Q^\bullet(\bar{k}, t(\dot{x}_1, \dots, \dot{x}_n) = s(\dot{x}_1, \dots, \dot{x}_n))$$

can be easily proved in S_2^1 by induction. We only have to transform it into a Σ_1^b formula, i.e. we have to find a polynomial bound to the length of the proof of $t(\dot{x}_1, \dots, \dot{x}_n) = s(\dot{x}_1, \dots, \dot{x}_n)$. By Lemma 5.21 the length of a proof of $I(\dot{x})$ is polynomial in the length of x . Next we find polynomial proofs of

- (e.1) $S(\text{numeral}(x_1)) = \text{numeral}(S(x_1)),$
- (e.2) $\text{numeral}(x_1) + \text{numeral}(x_2) = \text{numeral}(x_1 + x_2),$
- (e.3) $\text{numeral}(x_1) * \text{numeral}(x_2) = \text{numeral}(x_1 * x_2).$

Perhaps the best way to see a polynomial bound to these statements is the following. First construct proofs of $\dot{x}_i = t_i$ where t_i are terms of the form

$$\bar{2}^{j_1} + \dots + \bar{2}^{j_m}, \quad (j_1 > j_2 > \dots > j_m),$$

where $\bar{2}^k$ denotes $\underbrace{\bar{2} * (\bar{2} * \dots * \bar{2})}_{k \text{ times}}$. Think of t_i as a modified numeral. Then

follow the standard algorithms for $+$ and $*$ to obtain the proofs of (e.1–3) for the *modified numerals*. Finally transform the modified numerals back into original $\text{numeral}(S(x_1))$, $\text{numeral}(x_1 + x_2)$ and $\text{numeral}(x_1 * x_2)$ respectively. As is easily shown by induction, the maximal of these three bounds (the

bound for the length of the proof of (e.3) is asymptotically maximal) is a bound to the proof of the equality of general terms. To prove the inequality, take a suitable term u , prove

$$t(\dot{x}_1, \dots, \dot{x}_n) + S(u) = s(\dot{x}_1, \dots, \dot{x}_n),$$

or

$$s(\dot{x}_1, \dots, \dot{x}_n) + S(u) = t(\dot{x}_1, \dots, \dot{x}_n),$$

and use the fact that

$$Q \vdash I(x) \& I(y) \rightarrow x + S(y) \neq x.$$

□

5.23 Lemma. There exists k such that

$$S_2^1 \vdash (\forall y) RPr_Q^*(\bar{k}, (\forall x)(x \leq numeral^*(y+1) \equiv (x \leq numeral^*(y) \vee x = numeral^*(y+1))))).$$

Proof. We construct a cut I in Q such that

$$(e.4) \quad Q \vdash (\forall x, z)(I(z) \rightarrow (x \leq S(z) \equiv (x \leq z \vee x = S(z)))).$$

First take J defined by

$$J(x) \equiv_{df} 0 + x = x \& (\forall z)(S(z) + x = S(z+x)).$$

One can easily show that J is inductive. Now take I to be a *subcut* of J and verify (e.4). We shall work in S_2^1 . Let a number y be given. By Lemma 5.21 we have $Q \vdash I(numeral(y))$, hence by (e.4)

$$Q \vdash x \leq S(numeral(y)) \equiv (x \leq numeral(y) \vee x = S(numeral(y))).$$

By Lemma 5.22 $Q \vdash numeral(y+1) = S(numeral(y))$. Thus Q proves

$$(\forall x)(x \leq numeral(y+1) \equiv (x \leq numeral(y) \vee x = numeral(y+1))).$$

All that we have used were proofs of bounded cut-rank. □

For the following lemma, we shall fix some translation of L_2 into L_0 which determines an interpretation of *BASIC* on some cut in Q . Such a translation exists, since $IS_0 + \Omega_1$ is interpretable in Q by Theorem 5.12. We shall say that a formula ψ of L_0 is *essentially* Σ_1^b (resp. Π_1^b), if it is the translation of a Σ_1^b (resp. Π_1^b) formula of L_2 .

5.24 Lemma (formalized Σ completeness of Q).

(i) Let $\psi(x, y)$ be a bounded formula in L_0 . Then there exists k such that

$$I\Sigma_0 + Exp \vdash (\forall x)((\exists y)\psi(x, y) \rightarrow RPr_Q^\bullet(\bar{k}, (\exists y)\psi(\dot{x}, y))).$$

(ii) Let $\psi(x, y)$ be essentially Σ_1^b . Then there exists k such that

$$S_2^1 \vdash (\forall x)((\exists y)\psi(x, y) \rightarrow RPr_Q^\bullet(\bar{k}, (\exists y)\psi(\dot{x}, y))).$$

Proof. First we observe that in both cases it suffices to prove only

$$(\forall x)(\psi(x, y) \rightarrow RPr_Q^\bullet(\bar{k}, \psi(\dot{x}, \dot{y}))),$$

since $(\exists y)\psi(\dot{x}, y)$ will follow by an application of the rule $(\exists)$.

(i) The task of proving $\psi(\dot{x}, \dot{y})$ is similar to the task of deciding its truth: we shall prove $\psi(\dot{x}, \dot{y})$ by proving numerical instances of the open part of ψ . Thus we can bound the size of the numerals needed in this proof in the same way as in the proof of Theorem 5.4. The numerical instances of the open part of ψ follow from the numerical instances of atomic and negated atomic subformulae. The existence of bounded cut-rank proofs of atomic sentences with $=$ and their negations follows from Lemma 5.22. The atomic sentences with $\leq$ and their negations are reduced to atomic sentences with $=$ in the same way as the negations of atomic sentences with $=$ had been.

We must be a little careful when pasting the proofs of the numerical instances into a proof of a bounded sentence, since we need a proof of *bounded cut-rank*. A sentence of the form $(\exists y \leq t)\alpha(y)$ follows from some $\alpha(\bar{n})$, where n is less than or equal to the value of t . A proof of $(\forall y \leq t)\alpha(y)$ is constructed from the proofs of $\alpha(\bar{0}), \alpha(\bar{1}), \dots, \alpha(\bar{m})$, where m is the value of t , as follows. We prove successively $(\forall y \leq \bar{n})\alpha(y)$, for $n = 0, \dots, m$. For $n = 0$ it is an immediate consequence of $\alpha(\bar{0})$; to obtain it for $n + 1$ from a proof for n , use Lemma 5.23 and $\alpha(\bar{n} + 1)$. Finally use Lemma 5.22 to show $Q \vdash (\forall y \leq t)\alpha(y) \equiv (\forall y \leq \bar{m})\alpha(y)$.

(ii) The proof of (ii) is only a modification of the proof of (i) where we replace *bounded* by *sharply bounded*. We also need an extension of Lemma 5.23 to L_2 , instead of L_0 , and *BASIC*, instead of Q . Again the proof is essentially the same: translate the natural polynomial time algorithms for the evaluation of terms in L_2 into a proof using equalities which hold on a suitable cut in *BASIC*. We omit the details. $\square$

5.25 Lemma. Let $\psi(x)$ be a bounded formula and suppose

$$I\Sigma_0 + Exp \vdash (\forall x)\psi(x).$$

Then there exists k such that

$$I\Sigma_0 \vdash (\forall x)((\exists y)(y = 2_k^x) \rightarrow \psi(x)).$$

Proof. Suppose $I\Sigma_0$ does not prove $(\forall x)((\exists y)(y = 2_k^x) \rightarrow \psi(x))$, for any k . Then, by compactness, $I\Sigma_0 + \{(\exists y)(y = 2_k^c) \& \neg\psi(c) \mid k \in N\}$ (c a new constant) is consistent. Take a model M of this theory and an initial segment of it

$$K = \{a \in M \mid (\exists k \in N)(M \models a \leq 2_k^c)\}.$$

Then K is a model of $I\Sigma_0 + Exp$, but $K \models \neg\psi(c)$, because ψ is bounded. $\square$

Now we are ready for the theorems.

5.26 Theorem. Let $\psi(x)$ be a bounded formula. Then the following are equivalent:

- (1) There exists a cut $I(x)$ in Q such that

$$Q \vdash (\forall x)(I(x) \rightarrow \psi(x)).$$

- (2) There exists an interpretation of $Q + (\forall x)\psi(x)$ in Q .

- (3) $I\Sigma_0 + Exp \vdash (\forall x)\psi(x)$.

Proof. (1) $\Rightarrow$ (2) follows from the existence of shortenings of cuts closed under $+$ and $*$.

(2) $\Rightarrow$ (1) Let ι be an interpretation of $Q + (\forall x)\psi(x)$ in Q . We shall construct a cut $I(x)$ in $I\Sigma_0$ such that

$$I\Sigma_0 \vdash (\forall x)(I(x) \rightarrow \psi(x)).$$

This is sufficient, since $I\Sigma_0$ has an interpretation in Q determined by a cut. The cut $I(x)$ is defined by the formalization of the following formula: "There exists an isomorphism of the initial segment $[\bar{0}, x]$ onto an initial segment in the sense of interpretation ι ". Here we take intervals as structures with $=$, $\leq$, and $+$, $*$ as ternary relations. An isomorphism is a number which codes a sequence of pairs. We leave it to the reader to write down an explicit formula for I and check that it is a cut in $I\Sigma_0$. Let us only note that the extension of an isomorphism with domain $[\bar{0}, x]$ to an isomorphism with domain $[\bar{0}, S(x)]$ is obtained by adding to it a pair $(S(x), z)$, where z is the ι successor of the image of x . Such an extension of a sequence is possible in $I\Sigma_0$ (but not in Q). Now the fact that $(\forall x)\psi(x)$ holds in the sense of ι is transferred by the isomorphisms onto I .

- (1) $\Rightarrow$ (3) Suppose

$$Q \vdash (\forall x)(I(x) \rightarrow \psi(x)).$$

By Lemma 5.21

$$I\Sigma_0 + Exp \vdash (\forall x) RPr_Q^\bullet(\bar{k}, I(x)),$$

for some constant k . Hence

$$I\Sigma_0 + Exp \vdash (\forall x) RPr_Q^\bullet(\bar{k}, \psi(x)).$$

By Theorem 5.20, $I\Sigma_0 + Exp$ proves the consistency of Q with respect to proofs of cut-rank k . Thus we have

$$I\Sigma_0 + Exp \vdash (\forall x) \neg RPr_Q^\bullet(\bar{k}, \neg\psi(\dot{x})).$$

By Lemma 5.24(i)

$$I\Sigma_0 + Exp \vdash (\forall x)(\neg\psi(x) \rightarrow RPr_Q(\bar{m}, \neg\psi(\dot{x})))$$

for some m . We can assume that $k = m$ (otherwise take $\max(k, m)$ instead of k and m). Hence (3) follows from the last two statements.

(3) $\Rightarrow$ (1) Suppose $I\Sigma_0 + Exp \vdash (\forall x)\psi(x)$. By the lemma above we have some k such that

$$I\Sigma_0 \vdash (\forall x)((\exists y)(y = 2_k^x) \rightarrow \psi(x)).$$

Take cuts I, J in Q such that J determines an interpretation of $I\Sigma_0$ in Q and I is a subcut of J such that

$$Q \vdash I(x) \rightarrow (\exists y)(y = 2_k^x \ \& \ J(y)).$$

Then we have $Q \vdash (\forall x)(I(x) \rightarrow \psi(x))$. □

5.27 Theorem. Let $\psi(x)$ be a bounded formula in L_0 . Then the following are equivalent:

- (1) $I\Sigma_0 + Exp \vdash (\forall x)\psi(x)$.
- (2) For some k

$$S_2^1 \vdash (\forall x) RPr_Q^\bullet(\bar{k}, \psi(\dot{x})).$$

If $\psi(x)$ is the translation of a Π_1^b formula $\varphi(x)$ into L_0 , then (1) and (2) are equivalent to:

- (3) For some k

$$S_2^1 + RCon_Q^\bullet(\bar{k}) \vdash (\forall x)\varphi(x).$$

Proof. (1) $\Rightarrow$ (2) Suppose $I\Sigma_0 + Exp \vdash (\forall x)\psi(x)$, ψ bounded. Then by Theorem 5.26 we have $Q \vdash (\forall x)(I(x) \rightarrow \psi(x))$ for some cut I in Q . By Σ_1 completeness

$$S_2^1 \vdash RPr_Q^\bullet(\bar{j}, (\forall x)(I(x) \rightarrow \psi(x))),$$

for some j . By Lemma 5.21

$$S_2^1 \vdash (\forall x) RPr_Q^\bullet(\overline{m}, I(\dot{x}))$$

for some m . Take k to be the maximum of j and m , then we have (2).

(2) $\Rightarrow$ (1) By definition

$$(e.5) \quad S_2^1 + RCon_Q^\bullet(\overline{k}) \vdash (\forall x) RPr_Q^\bullet(\overline{k}, \psi(\dot{x})) \rightarrow (\forall x) \neg RPr_Q^\bullet(\overline{k}, \neg\psi(\dot{x})).$$

By Theorem 5.20, $RCon_Q^\bullet(\overline{k})$ is provable in $I\Sigma_0 + Exp$, thus (e.5) is provable in $I\Sigma_0 + Exp$. Since we assume (2), we have

$$I\Sigma_0 + Exp \vdash (\forall x) \neg RPr_Q^\bullet(\overline{k}, \neg\psi(\dot{x})).$$

By Lemma 5.24(i)

$$I\Sigma_0 + Exp \vdash (\forall x)(\neg RPr_Q^\bullet(\overline{k}, \neg\psi(\dot{x})) \rightarrow \psi(x)).$$

Hence (2) implies (1).

Now assume that ψ is the translation of a Π_1^b formula $\varphi(x)$ into L_0 .

(2) $\Rightarrow$ (3) By Lemma 5.24(ii) we have

$$S_2^1 \vdash (\forall x)(\neg RPr_Q^\bullet(\overline{k}, \neg\psi(\dot{x})) \rightarrow \psi(x)).$$

Then using (e.5) we get

$$S_2^1 + RCon_Q^\bullet(\overline{k}) \vdash (\forall x)\psi(x).$$

from (2). But $\varphi(x)$ is equivalent to $\psi(x)$ in S_2^1 , thus we have (3).

(3) $\Rightarrow$ (1) This is because $I\Sigma_0 + Exp$ "contains" S_2^1 and proves $RCon_Q^\bullet(\overline{k})$. $\square$

We would like to axiomatize Π_1 consequences of $I\Sigma_0 + Exp$ over, say, S_2^1 (put otherwise, we want to find some nice basis). The equivalence of (1) with (3) gives only a partial answer: since $RCon_Q^\bullet(\overline{k})$ can be written as a formula of the form $(\forall x)\psi(x)$ with ψ in Π_1^b , the set of sentences $RCon_Q^\bullet(\overline{k})$, $k = 0, 1, \dots$ axiomatizes the $\forall\Pi_1^b$ consequences of $I\Sigma_0 + Exp$.

(f) Incompleteness Theorems

It would be paradoxical if the incompleteness theorems did not hold in weak systems. Of course formalization of syntax in weaker theories is a more difficult task, therefore we cannot apply classical proofs of the second Gödel incompleteness theorem quite directly. It turns out that the second

incompleteness theorem can be even strengthened. A typical example is the unprovability of the consistency of Q in $I\Sigma_0 + Exp$. The results presented in this subsection will be corollaries of the theorems proved above and of the following theorem.

Note that in the following theorem (i) holds for every finitely axiomatized sequential theory containing Q . We shall not prove this general theorem here. For finite theories in the language of arithmetic extending $I\Sigma_1$, statement (i) follows from Theorem 3.20 (1) Chap. III. Statement (ii) generalizes Theorem 3.11, Chap. III, to weaker theories; however the assumption about the axiomatizability is stronger here. It is possible to use here the same assumption about the axiomatizability as in Chap. III, but we shall not prove this strengthening. Recall that $RCon_T^{\bullet J}(\bar{k})$ is

$$(\forall z)(I(z) \& c\text{-rank}(z) \leq \bar{k} \rightarrow \neg \text{Proof}_T^{\bullet}(z, \overline{0=1})).$$

5.28 Theorem. (i) For every k there exists a cut I in Q such that

$$Q \vdash RCon_Q^{\bullet J}(\bar{k});$$

(ii) Let T be a consistent theory containing Q and having a Σ_1^b axiomatization. Then for every cut J in T there exists m such that

$$\neg(T \vdash RCon_T^{\bullet J}(\bar{m})).$$

Proof. (i) By Theorem 5.20 (ii), for every k

$$I\Sigma_0 + Exp \vdash RCon_Q^{\bullet}(\bar{k}).$$

By Theorem 5.26, this implies the first statement.

(ii) The proof of this statement will be similar to the proof of Theorem 3.11, Chap. III. Roughly speaking, we only have to take m sufficiently large. Since this result plays a key role in the rest of this section, we shall prove it in detail.

Because of the shortening techniques, we may suppose that J has some additional properties. Thus we shall assume that J is a cut in Q and it determines an interpretation of S_2^1 in Q . In particular, it is provable in Q , hence also in T , that for any two sequences in J , their concatenation lies also in J .

We would like to use the provability conditions of 2.16, Chap. III, for $RPr_T(\bar{m}, x)$. We shall prove the first two. The third one, which is formalized *Modus Ponens*, is clearly false, since *Modus Ponens* increases the cut-rank. Thus we shall first take a suitable m , and then we shall follow the usual proof

of the second incompleteness theorem and check that the third condition is actually applied only to formulae of rank at most m . We shall divide the proof into several claims.

Claim 1. For every sentence φ , if $T \vdash \varphi$, then $T \vdash \text{CFPr}_T^{\bullet J}(\overline{\varphi})$.

Proof. Suppose $T \vdash \varphi$. By cut-elimination we have a cut-free proof of φ in T . By the Σ -completeness of Q , $Q \vdash \text{CFPr}_T^{\bullet J}(\overline{\varphi})$. Since J determines an interpretation of Q in T , we get the conclusion of the claim. $\square$

Claim 2. There exists k such that

$$T \vdash \text{RPr}_T^{\bullet J}(x, y) \rightarrow \text{RPr}_T^{\bullet J}(\overline{k, \text{RPr}_T^{\bullet J}(\dot{x}, \dot{y})}).$$

Proof. Let us first write down $\text{RPr}_T^{\bullet J}(x, y)$ more explicitly:

$$(\exists d)(J(d) \& J(x) \& J(y) \& c\text{-rank}(d) \leq x \& \text{Proof}_T(d, y)).$$

By Lemma 5.24, for some k_1

$$\begin{aligned} S_2^1 \vdash \text{rank}(d) \leq x \& \text{Proof}_T^{\bullet}(d, y) \rightarrow \\ \rightarrow \text{RPr}_Q^{\bullet}(\overline{k_1, \text{rank}(\dot{d}) \leq x \& \text{Proof}_T^{\bullet}(\dot{d}, \dot{y})}). \end{aligned}$$

By Lemma 5.21, for some k_2

$$S_2^1 \vdash \text{RPr}_Q^{\bullet}(\overline{k_2, J(\dot{d}) \& J(\dot{x}) \& J(\dot{y})}).$$

To obtain $\text{RPr}_T^{\bullet J}(\dot{x}, \dot{y})$ we need only the rules for $\&$ and $\exists$ which do not increase the cut-rank. Hence for $k = \max(k_1, k_2)$

$$S_2^1 \vdash \text{RPr}_T^{\bullet}(x, y) \rightarrow \text{RPr}_Q^{\bullet}(\overline{k, \text{RPr}_T^{\bullet J}(\dot{x}, \dot{y})}).$$

Here we can replace Q by T , since Q is contained in T . As $T \vdash (S_2^1)^J$, we get

$$T \vdash (\text{RPr}_T^{\bullet}(x, y) \rightarrow \text{RPr}_T^{\bullet}(k, \overline{\text{RPr}_T^{\bullet J}(\dot{x}, \dot{y})}))^J,$$

which is in fact the statement of the claim. $\square$

By 2.1, Chap. III, we have some $\delta(x)$ such that

$$T \vdash \delta(x) \equiv \neg \text{RPr}_T^{\bullet J}(x, \overline{\delta(\dot{x})}).$$

$\overline{\delta(x)}$ is the numeral of the formula $\delta(\text{numeral}(x))$. Now we can choose m . Let m be such that

- (1) $m \geq k$, where k is from Claim 2;
- (2) $m \geq \text{rank}(RPr_T^{\bullet J}(x, y))$;
- (3) $m \geq \text{rank}(\delta(x))$.

Note that we thus have also $m \geq \text{rank}(RPr_T^{\bullet J}(\overline{m}, y))$ and $m \geq \text{rank}(\delta(\overline{m}))$. Let us define also

$$\begin{aligned}\Delta &=_{df} \delta(\overline{m}); \\ \alpha &=_{df} RPr_T^{\bullet J}(\overline{m}, \overline{\Delta}).\end{aligned}$$

Claim 3. $T \vdash \Delta \equiv \neg RPr_T^{\bullet J}(\overline{m}, \overline{\Delta})$.

Proof. By definition we have

$$T \vdash \delta(\overline{m}) \equiv \neg RPr_T^{\bullet J}(\overline{m}, \overline{\delta(\overline{m})}).$$

Thus we only need

$$T \vdash \overline{\delta(\overline{m})} = \overline{\delta(\overline{m})},$$

which follows from the Σ -completeness of Q , hence of T . $\square$

Claim 4. $\neg(T \vdash \Delta)$.

Proof. Suppose $T \vdash \Delta$. Then, by Claim 1, $T \vdash CFP_r_T^{\bullet J}(\overline{\Delta})$, whence $T \vdash RPr_T^{\bullet J}(\overline{m}, \overline{\Delta})$. By Claim 3 this means that T is inconsistent, which is a contradiction. $\square$

Claim 5. $T \vdash RPr_T^{\bullet J}(\overline{m}, \overline{\alpha}) \& RPr_T^{\bullet J}(\overline{m}, \neg \overline{\alpha}) \rightarrow \neg RCon_T^{\bullet J}(\overline{m})$.

Proof. If α and $\neg \alpha$ are derivable using proofs of cut-rank of at most m , then so are $\alpha \vee \overline{0} = \overline{1}$ and $\neg \alpha \vee \overline{0} = \overline{1}$ (add $\overline{0} = \overline{1}$ to each proof line as a side formula). Since $m \geq \text{rank}(RPr_T^{\bullet J}(x, y)) = \text{rank}(\alpha)$, we get a proof of $\overline{0} = \overline{1}$ by taking the concatenation of these proofs and applying the cut rule. The closure properties of J ensure that such a proof will be also in J . $\square$

Claim 6. $T \vdash RPr_T^{\bullet J}(\overline{m}, \overline{\Delta}) \rightarrow RPr_T^{\bullet J}(\overline{m}, \overline{\alpha})$.

Proof. By Claim 2

$$T \vdash RPr_T^{\bullet J}(\overline{m}, \overline{\Delta}) \rightarrow RPr_T^{\bullet J}(\overline{m}, \overline{RPr_T^{\bullet J}(\overline{m}, \overline{\Delta})}).$$

By the definition of α and Σ -completeness

$$T \vdash \overline{RPr_T^{\bullet J}(\overline{m}, \overline{\Delta})} = \overline{\alpha}.$$

$\square$

Claim 7. $T \vdash RPr_T^{\bullet J}(\overline{m}, \overline{\Delta}) \rightarrow RPr_T^{\bullet J}(\overline{m}, \neg \overline{\alpha})$.

Proof. By Claim 3 and the definition of α , $T \vdash \Delta \rightarrow \neg \alpha$. By Claim 1, this implies $T \vdash RPr_T^{\bullet J}(\overline{m}, \overline{\Delta} \rightarrow \neg \overline{\alpha})$. If we have proofs of Δ and $\Delta \rightarrow \neg \alpha$ of cut-rank m , we can easily combine them into a proof of cut-rank m of $\neg \alpha$, because $m \geq rank(\Delta)$. Again, the closure properties of J ensure that this proof is in J too. $\square$

Now we can finish the proof of part (ii) of the theorem. By Claims 3, 6, 7 and 5, we have in T

$$\neg \Delta \rightarrow RPr_T^{\bullet J}(\overline{m}, \overline{\Delta}) \rightarrow (RPr_T^{\bullet J}(\overline{m}, \overline{\alpha}) \& RPr_T^{\bullet J}(\overline{m}, \neg \overline{\alpha})) \rightarrow \neg RCon_T^{\bullet J}(\overline{m}),$$

i.e. $T \vdash RCon_T^{\bullet J}(\overline{m}) \rightarrow \Delta$. Hence, by Claim 4, $RCon_T^{\bullet J}(\overline{m})$ is not provable in T . $\square$

5.29 Corollary. Con_Q is not provable in $I\Sigma_0 + Exp$.

Proof. By the theorem above, Q does not prove $Con_Q^{\bullet I}$ for any cut I . Hence by Theorem 5.26 $I\Sigma_0 + Exp$ does not prove $Con_Q^{\bullet}$. $\square$

5.30 Corollary. Cut elimination (unrestricted), and Herbrand's Theorem are not provable in $I\Sigma_0 + Exp$.

Proof. By Theorem 5.20 (ii) $I\Sigma_0 + Exp$ proves cut-free and Herbrand consistency of Q . If we had cut-elimination or Herbrand's Theorem in $I\Sigma_0 + Exp$, we could reduce the ordinary consistency of Q to the restricted one. $\square$

5.31 Theorem. $I\Sigma_0 + Exp + Con_{I\Sigma_0}^{\bullet}$ does not prove $Con_{I\Sigma_0 + Exp}^{\bullet}$.

Proof. Let J be the cut in $I\Sigma_0 + Exp$ defined by

$$J(x) \equiv_{df} (\exists y)(y = 2_x^0).$$

We shall show that

$$(f.1) \quad I\Sigma_0 + Exp \vdash Con_J^{\bullet}.$$

Suppose not. Let us work in $I\Sigma_0 + Exp$, let d be a proof of a contradiction in J . Then d has cut-rank at most $lh(d)$. We have

$$2_{3*lh(d)}^{lh(d)} \leq 2_d^0,$$

since d must be nonstandard and d is exponentially larger than $lh(d)$; hence such a number exists. Thus we can apply cut elimination to obtain a cut-free

proof of a contradiction in Q (apply Lemma 5.15 $lh(d)$ -times and use the inequality

$$2^{2^{2 \cdot lh(d)}} \leq 2^{2^{lh(d)}}).$$

But this is not possible by 5.20 (ii).

Now suppose the statement of the theorem is false. Thus

$$I\Sigma_0 + Exp \vdash Con_{I\Sigma_0}^\bullet \rightarrow Con_{I\Sigma_0 + Exp}^\bullet.$$

Since $Con_{I\Sigma_0}$ and Con_Q are equivalent in $I\Sigma_0 + Exp$, we have

$$I\Sigma_0 + Exp \vdash Con_Q^\bullet \rightarrow Con_{I\Sigma_0 + Exp}^\bullet.$$

Using an obvious Lemma 5.25, we can deduce that provably in $I\Sigma_0 + Exp$, a proof d of a contradiction in $I\Sigma_0 + Exp$ can be transformed into a proof d' of a contradiction in Q such that $d' \leq 2_k^d$ for a suitable standard k . Let I be a subcut of J such that

$$I\Sigma_0 + Exp \vdash I(x) \rightarrow J(2_k^x).$$

Hence a proof d of a contradiction in $I\Sigma_0 + Exp$ such that d is in I can be transformed into a proof d' of a contradiction in Q such that d' is in J . Thus, by (f.1) no such proof can exist, i.e. $I\Sigma_0 + Exp$ proves $Con_{I\Sigma_0 + Exp}^\bullet$. But this is in contradiction to Theorem 5.28 (ii). $\square$

5.32 Corollary. $I\Sigma_0 + Exp$ is not interpretable in $I\Sigma_0$.

Proof. Suppose $I\Sigma_0 + Exp$ is interpretable in $I\Sigma_0$. It is well-known that if T is interpretable in S and S is consistent, then T is consistent too. This simple theorem can be easily formalized in $I\Sigma_0 + Exp$, if T is finitely axiomatized. Since $I\Sigma_0 + Exp$ is finitely axiomatizable (Theorem 5.6), as one can check provably in $I\Sigma_0 + Exp$, we get

$$I\Sigma_0 + Exp \vdash Con_{I\Sigma_0}^\bullet \rightarrow Con_{I\Sigma_0 + Exp}^\bullet,$$

which contradicts Theorem 5.31. $\square$

5.33 Theorem. For every k there exists m such that $RCon_Q^\bullet(\overline{m})$ is not provable in $I\Sigma_0 + \Omega_n + RCon_Q^\bullet(\overline{k})$ for any n .

Proof. We shall use Theorem 5.28 again. Let I be a cut in Q such that

$$Q \vdash RCon_Q^\bullet(\overline{k}).$$

We can suppose that I determines an interpretation of $I\Sigma_0$. Let J be a subcut of I such that

$$Q \vdash J(x) \rightarrow I(2^{2x}).$$

We know that, for some m , $Q + \neg \text{RCon}_Q^{\bullet J}(\bar{m})$ is consistent. Let M be a model of this theory. Let K be its initial segment defined by

$$K = \{a \in M; (\exists b \in J^M)(\exists i \in N)(a \leq \omega_i(b))\}.$$

First we shall show that K is closed under all functions ω_n . Let $a \in K$; without loss of generality we can suppose that a is nonstandard. Then, for some nonstandard $b \in J^M$, $a \leq \omega_i(b)$. Let n be given. Then

$$\begin{aligned} \omega_n(a) &\leq \omega_n(\omega_i(b)) \leq \omega_{\max(n,i)}(\omega_{\max(n,i)}(b)) \\ &\leq \omega_{\max(n,i)+1}(b) \in K. \end{aligned}$$

To prove the last inequality, use induction on j to show that

$$\omega_j(\omega_j(x)) \leq \omega_{j+1}(x),$$

for every sufficiently large x (hence for every nonstandard x). Now we need that $K \subseteq I^M$. This follows from the definition of the cuts and from the following inequality, which is also easily provable by induction on j :

$$\omega_j(x) \leq 2^{2^x}, \quad \text{for } x \text{ sufficiently large.}$$

Since K is contained in I^M , which is a model of $I\Sigma_0$, K is a model of $I\Sigma_0 + \omega_n$, for every n . Since $Q \vdash \text{RCon}_Q^{\bullet I}(\bar{k})$ and $M \models \neg \text{RCon}_Q^{\bullet J}(\bar{m})$, we have

$$K \models \text{RCon}_Q^{\bullet}(\bar{k}) \ \& \ \neg \text{RCon}_Q^{\bullet}(\bar{m}),$$

which finishes the proof. $\square$

This theorem gives us another piece of important information about $\forall \Pi_1^b$ consequences of $I\Sigma_0 + \text{Exp}$: they are not “finitely based”. Further we have this consequence.

5.34 Corollary. $I\Sigma_0 + \text{Exp}$ is not Π_1 conservative over $I\Sigma_0 + \{\Omega_n; n \in N\}$.

In fact this is all that we know about Π_1 conservativity between systems of bounded arithmetic. In particular it seems plausible that $I\Sigma_0 + \Omega_1$ is not Π_1 conservative over $I\Sigma_0$, but we cannot even prove that $I\Sigma_0 + \{\Omega_n; n \in N\}$ is not Π_1 conservative over $I\Sigma_0$. A Π_1 sentence which might separate $I\Sigma_0 + \Omega_1$ from $I\Sigma_0$ is the so-called Bertrand’s postulate. It is the theorem (due to Sylvester) which says that there is a prime between n and $2n$, for every n .

This sentence is provable in $I\Sigma_0 + \Omega_1$, but no proof of it has been found in $I\Sigma_0$. Whether such a proof exists is an important open problem.

(g) On the Limited Use of Exponentiation

If we are interested only in Π_1 sentences, then it seems that all such sentences which are also truly mathematical results are provable in $I\Sigma_0 + Exp$ (we consider sentences like $Con(Q)$ to be uninteresting for an ordinary mathematician). In fact the proofs use exponentiation at most once. This means that for instance if we construct some graph, we may consider the set of all graphs of a given size, but usually we do not consider the set of subsets, etc. Put otherwise, we use the finite power set operation only once. This phenomenon is worthwhile to formalize, but it cannot be done directly in first order logic. In first order theory we have either to accept axiom Exp , and then we have also all finite iterations of 2^x , or we do not have exponentiation at all. A possible approach is to use higher order bounded arithmetics. We shall use a different approach. Our aim is, roughly speaking, to show that there is an infinite hierarchy according to the number of applications of the exponentiation axiom.

So what does it really mean that $(\forall x)\varphi(x)$ is provable using exponentiation only once? This means that we prove $\varphi(x)$ assuming that 2^x exists; formally

$$I\Sigma_0 + \Omega_1 \vdash (\forall x)((\exists y)(y = 2^x) \rightarrow \varphi(x)).$$

It is convenient to consider $I\Sigma_0 + \Omega_1$ instead of $I\Sigma_0$, since thus we obtain a concept which is more robust: in $I\Sigma_0 + \Omega_1$, 2^x exists iff 2^{x^2} exists, etc. There is an equivalent and shorter way to express that $(\forall x)\varphi(x)$ is provable using exponentiation only once, which is

$$I\Sigma_0 + \Omega_1 \vdash (\forall x)\varphi(|x|).$$

Similarly, k applications correspond to the provability of $(\forall x)\varphi(|x|^{(k)})$, where $|x|^{(k)}$ denotes the k times iterated function $x \mapsto |x|$.

Finally let us also mention a model theoretical characterization of provability using exponentiation only once. Let $I(x)$ be defined by

$$I(x) \equiv_{df} (\exists y)(y = 2^x).$$

Let K be the class of models of the form

$$K = \{a \in M; M \models I\Sigma_0 + \Omega_1 \ \& \ M \models I(a)\},$$

with the operations inherited from M . Then the sentences in question are just the sentences true in all models of K .

Let $\psi(x, y)$ be a bounded formula such that for every m

$$I\Sigma_0 + \Omega_1 \vdash (\forall y)\psi(\overline{m}, y) \equiv RCon_Q^\bullet(\overline{m}).$$

5.35 Lemma. For every k there exists m such that $(\forall y)\psi(\overline{m}, |y|^{(k)})$ is not provable in $I\Sigma_0 + \Omega_1$.

Proof. Let $I(x)$ be a cut in Q such that I determines an interpretation of $I\Sigma_0 + \Omega_1$. Let $J(x)$ be a subcut of $I(x)$ such that

$$Q \vdash J(x) \rightarrow I(2_k^x).$$

By Theorem 5.28, there exists m such that

$$\neg(Q \vdash RCon_T^{\bullet J}(\overline{m})).$$

Now suppose that $(\forall y)\psi(\overline{m}, |y|^{(k)})$ is provable in $I\Sigma_0 + \Omega_1$. Then

$$Q \vdash ((\forall y)\psi(\overline{m}, |y|^{(k)}))^I,$$

hence, by the construction of J ,

$$Q \vdash ((\forall y)\psi(\overline{m}, y))^J,$$

which is a contradiction. □

5.36 Theorem. For every k there exists a bounded formula $\varphi(x)$ such that

$$I\Sigma_0 + \Omega_1 \vdash (\forall x)\varphi(|x|^{(k+1)}),$$

but

$$\neg(I\Sigma_0 + \Omega_1 \vdash (\forall x)\varphi(|x|^{(k)})).$$

Proof. Let m be such that $(\forall y)\psi(\overline{m}, |y|^{(k)})$ is not provable in $I\Sigma_0 + \Omega_1$. Since $RCon_Q^\bullet(\overline{m})$, i.e. $(\forall y)\psi(\overline{m}, y)$, is provable in $I\Sigma_0 + Exp$, it must be provable using exponentiation a finite number of times, see Lemma 5.25. Thus, for some n ,

$$I\Sigma_0 + \Omega_1 \vdash (\forall x)\psi(\overline{m}, |x|^{(n)}).$$

Hence for some i , $k \leq i < n$,

$$I\Sigma_0 + \Omega_1 \vdash (\forall x)\psi(\overline{m}, |x|^{(i+1)}),$$

and

$$\neg(I\Sigma_0 + \Omega_1 \vdash (\forall x)\psi(\overline{m}, |x|^{(i)})).$$

Take $\varphi(x)$ to be $\psi(\overline{m}, |x|^{(i-k)})$, then the theorem follows. □

The instances of the schema $PHP\Sigma_0$ are Π_1 and they are provable using one exponentiation. Actually, most of the proofs that require exponentiation can be reduced to $PHP\Sigma_0$. Still it seems that $PHP\Sigma_0$ is not sufficient to prove all sentences derivable using one exponentiation. We are not able to derive $PHP\Sigma_0$ in $I\Sigma_0 + \Omega_1$, so it is possible that some instances could be used as more natural formulae in the theorem above for $k = 0$. However we do not know which. The point is that the known mathematical statements which can be derived using $PHP\Sigma_0$ can be derived also from a weaker version of PHP which is already provable in $I\Sigma_0 + \Omega_1$ (Bertrand's postulate is an example).

The related problems of whether $I\Sigma_0 + \Omega_i$ is Π_1 conservative over $I\Sigma_0 + \Omega_j$, for $j < i$, are still open. They are typical in the sense that they are related to open problems in complexity theory. For instance, the question whether $I\Sigma_0 + \Omega_1$ is Π_1 conservative over $I\Sigma_0$ seems to be formally related to the open problem whether $PH = LinH$ (though we are not able to prove any relation). New techniques will need to be developed in order to solve such problems. Solving such problems in Bounded Arithmetic might be the first step in solving the corresponding persistent problems in complexity theory.

Bibliographical Remarks and Further Reading

Preliminaries

We try to indicate for each important result or notion its author and the corresponding publication, and possibly also reference to another work where the result is presented. Our aim is to be as precise as possible; on the other hand, these remarks are not intended to be a complete historical source and they serve only for orientation. This concerns mainly remarks on old results. The reader interested in deeper investigation of origins of the metamathematics of arithmetic is referred to source books: Gödel's Collected Works I [Gödel 86], the books *From Frege to Gödel* [van Heijenoort 87] and *The Undecidable* [Davis 65]. We find also [Meschkowski 81] very informative. [Smoryński 91 – Logical] contains rather detailed historical information.

The material covered in the Preliminaries belongs to classics and can be found in standard monographs on mathematical logic, notably [Shoenfield 67], [Mendelson 64], [Bell-Machover 77] and others; note that the syntax of first order predicate logic is systematically developed in [Hilbert-Ackermann 28], largely in the style which is still in use. They formulate the problem of completeness; Gödel presents his solution in his dissertation (1930) published as [Gödel 31 – Monatsh.]. [Skolem 20] contains a proof of what we now call the downward Löwenheim-Skolem theorem, using what we call Skolemization. The above works seem to deal with validity and satisfiability as intuitively clear notions without trying to formalize them. Tarski's paper ([Tarski 33] in Polish; a German version is [Tarski 36], for an English version see [Tarski 56]) presents conditions defining the satisfaction of a compound formula from the satisfaction of its components – these are our “Tarski's truths conditions” or “Tarski's satisfaction conditions” (for a detailed analysis of Tarski's approach see [Hodges 85]).

Herbrand's theorem is contained in his thesis published as [Herbrand 30]; as commented in [van Heijenoort 67] p. 526, Herbrand considered his theorem to be a more precise statement of the well-known Löwenheim-Skolem theorem. From the above mentioned monographs, only [Shoenfield 67] elaborates on Herbrand's theorem.

Origins of first order arithmetic were described in the Introduction; let us add some details. Arithmetical hierarchy was introduced in [Kleene 43] and [Mostowski 47]; the notation Σ_n , Π_n , Δ_n is due to [Addison 58] and [Mostowski 59] (cf. Kleene's introductions in [Gödel 86]). Definition of functions from other functions by primitive recursion were known to Dedekind, Skolem, Hilbert and Ackermann. Gödel introduces in [Gödel 31 – Monatsh.] a class of functions that he calls recursive; in our present terminology, these are just primitive recursive functions. For origins of general recursive functions see [Péter 34], [Kleene 36]; monographs on recursive functions and recursion theory include [Rogers 67], [Soare 87]. Gödel introduces and uses coding of finite sequences of natural numbers by natural numbers [Gödel 31] and arithmetization of syntax. We shall comment more on arithmetization in remarks on Chaps. I and III.

Matiyasevič's theorem is a final step of a long research into Hilbert's tenth problem in which J. Robinson, M. Davis and H. Putnam played prominent roles; besides Matiyasevič's papers, see [Davis 73 Am. Math. Monthly] for history and a detailed proof. Finally let us note that the decidability of the set of all sentences of the [language of arithmetic without multiplication] true in N mentioned in 0.49 is proved in [Presburger 30]. Similarly, [Skolem 30] proved that the theory of N in the language with multiplication and equality is decidable; investigation of the theory of multiplication has been continued, see [Nadel 80], [Cegielsky 81].

Chapter I

(1) The theory Q was introduced by R.M. Robinson and is developed in [Tarski, Mostowski, Robinson 53]. First-order arithmetic is mentioned in Gödel's postscript [Gödel 32 - Ergebnisse] to his main paper on incompleteness [Gödel 31]; [Skolem 23] develops a first order quantifier free system known presently as primitive recursive arithmetic. His paper is devoted to "positive results": no metamathematical results are obtained. Coming back to the question why first order arithmetic with induction for all formulas is commonly called Peano arithmetic, let us mention that the first occurrence of this term known to us is in [Tarski-Mostowski-Robinson 53] where the authors say that the theory in question "is known as Peano arithmetic". [Kleene 52] studies the same system (or an equivalent system); in an introduction in [Gödel 86] Kleene says that his system "can be described as Peano arithmetic PA ". Kleene stresses the necessity of distinguishing PA from the original Peano's axioms that are second-order.

In developing I_{open} we freely follow [Mendelson 64]; but we leave I_{open} rather soon. It should be mentioned that there are several papers devoted to the metamathematics of I_{open} , notably [Shepherdson 65], [Wilkie 77], [van der Dries 80], [Adamowicz 86, 87] and others.

IS_0 and related systems of bounded arithmetics are studied extensively in Chap. V (see bibliographical remarks to that chapter). In developing IS_1 we partially anticipate what will be done in Sect. 2 for arbitrary IS_n ($n > 1$); the main task - arithmetization - is done in a rather standard way, usual in presenting Gödel's results on PA , and it is seen that IS_1 suffices to prove all necessary things. (The auxiliary coding of ω -sequences follows [Shoenfield 67].) Partial truth definitions are of great importance; they were studied by [Montague 59] ([Schütte 60] has a similar notion). [Kreisel-Lévy 68] is an important paper dealing with partial truth definitions; we shall mention it in remarks to Chap. III.

(2) The main starting point for the study of theories IS_n and $B\Sigma_{n+1}$ is the paper [Paris-Kirby 78]; this was preceded by [Parsons 70, 72] where some important partial results were obtained. The paper by Paris-Kirby contains among other things the equivalence of IS_n , III_n , LS_n , LII_n (Th. 2.4) and the implications $IS_{n+1} \Rightarrow B\Sigma_{n+1} \Rightarrow IS_n$, $B\Sigma_{n+1} \Leftrightarrow BII_n$ (Th. 2.5). The fact (2.4) that $IS_n \Rightarrow IS_0(\Sigma_n)$ appears in [Clote 85 - Caracas] (with a remark that Paris and Kossak independently proved the same result); the paper also contains a proof of the implication $B\Sigma_{n+1} \Rightarrow L\Delta_{n+1} \Rightarrow I\Delta_{n+1}$ (2.5). The implication $L\Delta_{n+1} \Rightarrow B\Sigma_{n+1}$ was proved by R.O. Gandy (unpublished). Note that H. Friedman circulated a preprint (on fragments of Peano arithmetic, before 1985) in which he claimed $B\Sigma_{n+1}$ and $I\Delta_{n+1}$ to be equivalent; but it still seems to be open whether $I\Delta_{n+1}$ proves $B\Sigma_{n+1}$ (or, equivalently, $L\Delta_{n+1}$). The fact that in $B\Sigma_{n+1}$, Σ_n and Π_n formulas are closed under bounded quantification is due to [Parsons 70].

The notion "piecewise coded" is studied in [Clote 85 - Caracas] and was introduced by Kossak. Theorem 2.7 appears in the same paper. Concerning theorem 2.23: The equivalence $S\Sigma_{n+1} \Leftrightarrow S\Sigma_n \Leftrightarrow IS_n$ is in [Clote 85 - Caracas], properties of PHP in [Dimitracopoulos-Paris 86], properties of the axioms of regularity in [Mills-Paris 84]. 2.24 (1) seems to be a folklore, properties of local and partial approximations (T and P) appear in [Hájek-Paris 86]. Theorem 2.25 is due to V. Švejdar (unpublished).

Finite axiomatizability of $I\Sigma_n$, $B\Sigma_{n+1}$ ($n \geq 1$) is an easy consequence of the existence of partial satisfactions; it is difficult to say who first observed it.

Relativized satisfaction (for $\Sigma_n^*(X)$ -formulas, X being a set) is introduced and investigated in [Hájek-Kučera 89]. [Clote 85 – Oberwolfach] works with $\Sigma_n(X)$ -definable subsets of a model M of $I\Sigma_1$, X being a subset of M ; but he does not formalize relativized satisfaction inside $I\Sigma_1$. The development of a theory of low Δ_{n+1} sets and sets of the order type of the universe (2.63–2.71) follows Clote's work but is made *inside* the theories $B\Sigma_{n+1}$, not from outside using models. The strengthening to $I\Sigma_n$ and replacement of (low) Δ_{n+1} by (low) $\Sigma_0^*(\Sigma_n)$ is due to [Hájek-Kučera 89]. We shall comment more on low sets in remarks to Sect. (3).

Finally, arithmetic with a top is studied in [Dimitracopoulos-Paris 82] and [Cegielsky-McAloon-Wilmers 82]. In 1.2 we have presented only rather elementary facts on this arithmetic.

(3) The limit theorem is a classical result of recursive theory due to [Putnam 65]. [Clote 85 – Caracas] proves the limit theorem for Δ_2 sets in $I\Sigma_1$ and for Δ_2 functions in $B\Sigma_2$. The limit theorem for Δ_2 functions in $B\Sigma_1$ was proved by Švejdar (3.2 – unpublished).

The low basis theorem is a recursion-theoretic version of the celebrated König's lemma [König 27] and is due to [Jockusch-Soare 72 – TAMS]. In the recursion-theoretic language, a degree a is low if $a' = 0'$ (see e.g. [Soare 87]). It is easy to show that the Turing degree of a set X is $0'$ iff $X \in \Delta_2$ and that $dg(X)$ is low iff X is low Δ_2 in our meaning. The low basis theorem says that if T is a recursive dyadic infinite tree then T has an infinite low branch. [Clote 85] shows that this is true in each $M \models B\Sigma_2$; we follow his proof, working inside $B\Sigma_2$. The improvement to $I\Sigma_1$ and low $\Sigma_0^*(\Sigma_1)$ is in [Hájek-Kučera 88]. 3.24 was proved by Paris (unpublished). [Gaifman 70] was the first to realise that Matiyasevič's theorem is provable in PA ; Dimitracopoulos proved Matiyasevič's theorem in $I\Sigma_0(\exp)$; cf. [Gaifman-Dimitracopoulos 82].

(4) As we already said in comments to the Preliminaries, the arithmetization of logic goes back to Gödel. The first systematic elaboration of logic inside first order arithmetic (arithmetization of syntax as well as semantics including the Arithmetized Completeness Theorem) is contained in [Hilbert-Bernays 34, 39]. (Clearly, they do not have the Low Arithmetized Completeness Theorem.) Checking that (some) usual proofs of logical facts formalize in $I\Sigma_1$ is more or less immediate (disregarding some difficulties as shown in the proof of 4.10). The Arithmetized Completeness Theorem can be stated as follows. PA proves that each consistent Δ_1 theory has a full Δ_2 model. Partial improvements were obtained by [Paris 81] and [McAloon 78 – TAMS]; our Low Arithmetized Completeness Theorem is more general.

Theorem 4.33 as well as Corollary 4.34 is rather important; it is closely related to the results of [Leivant 81]. Note that 4.34 (1) ($I\Sigma_{k+1} \vdash Con^*(I\Sigma_k)$) can be obtained by proof-theoretical methods, see [Takeuti 75]. (We shall comment on proof theory in remarks to Chap. III).

Finally, Theorem 4.37 is extracted from [Paris 80] and will be used in Chap. II where results of Paris's paper will be presented in detail.

Chapter II

As mentioned in the introduction to the present chapter, the interest in developing finite combinatorics in PA and its fragments arose from metamathematical investigations by Paris and his school. Paris's pioneering contribution is twofold: he produced a mathematically well-understood combinatorial Π_2 -theorem which is true but unprovable in PA , and he showed the unprovability by model-theoretic means, not by self-reference (as Gödel did when considering his unprovable true sentences). The original formulation of the principle now known as the Paris-Harrington principle was different from that presented here and did not use the notion of a large set. It circulated in preprints and may be found in

[Paris 78], [McAloon 79]. Harrington contributed by reformulating the principle into its elegant form using the notion of a (relatively) large finite set and showing the principle to be equivalent to $\text{Con}^*(PA + \text{Tr}(II_1))$ (cf. [Paris-Harrington 77]); the last equivalence was independently proved by [McAloon 80 – rapports]. Investigation of instances of the Paris-Harrington principle and their relation to fragments of PA (as well as of instances of the principle of ordinal-large intervals, see below) is the main content of Paris's papers [Paris 80], [Paris 81]. Positive results on fragments contained in these two papers form the main topic of the present chapter.

(1) Theorem 1.9 in particular is due to Paris. But we do not present it following Paris (since he obtained the provabilities in question by indirect, model-theoretical means). We follow and elaborate proofs of [Clote 85 – Oberwolfach] based on the Low Basis Theorem. In particular, Theorems 1.5, 1.6, 1.7 are due to Clote. Clote also reasons model-theoretically but his arguments easily yield direct proofs in fragments. Furthermore, we improve Clote's results by using our stronger version of Low Basic Theorem. 1.10 is folklore.

(2) Here we elaborate the proof of [Paris 81]; the subsection (b) on combinatorics relies on [Harrington-Paris 77].

(3) The history of what we call the Schwichtenberg-Wainer hierarchy or fast growing hierarchy is as follows (see [Buchholz-Wainer 87]): [Kreisel 52 – non fin.] showed that the functions provably recursive in PA can all be defined by recursions on certain well orderings of type $< \epsilon_0$. Later [Schwichtenberg 71] and [Wainer 70, 72] independently generalized earlier results of [Grzegorzczak 53] and [Robbin 65] to show that Kreisel's functions can be characterized by means of the present fast growing hierarchy below ϵ_0 . As further reference we mention [Löb-Wainer 70], [Schwichtenberg 77], [Rose 84], [Buchholz 84].

[Ketonen and Solovay 81] related this hierarchy to the Paris-Harrington principle and established, using purely combinatorial means, sharp upper and lower bounds to the function

$$\sigma(n) = \min\{a \mid [0, a] \rightarrow^* (n+1)^n_n\};$$

from this, they reproved Paris and Harrington's result. In their paper, Ketonen and Solovay introduced and studied the notion of an α -large finite set (due originally to Ketonen), α being an ordinal. It follows from their investigations that the principle

$$(W) \quad (\forall \alpha \leq \epsilon_0)(\forall x)(\exists y)([x, y] \text{ is } \alpha\text{-large})$$

is (meaningful and) unprovable in PA . [Paris 80, 81] introduced and investigated instances of (W) and related them to fragments of PA . We freely follow Paris's papers using [Ketonen-Solovay 81]. Theorem 3.18 does not occur explicitly in [Paris 81]; it can be found e.g. in [Takeuti 75], see also [Kurata 86].

There are various other important combinatorial principles that can be analyzed with respect to fragments, notably: a principle proposed independently by Pudlák (original paper unpublished, see [Hájek-Paris 86]), Kanamori-McAloon's principle [Kanamori-McAloon 87], principles due to [Clote-McAloon 83] and possibly others.

Chapter III

(1) The informal notion of an interpretation of a theory in another one appears to be rather old; but we did not attempt to identify particular references. The first work dealing with interpretations in connection with systems of first order arithmetic is [Tarski-Mostowski-Robinson 53]; an old paper is [Montague 57]. Feferman's fundamental paper [Feferman 60] deals also with interpretability; [Montague 63] contains a model-theoretic characterization of interpretability (cf. also [Hájek 66]).

Similarly, we present no information about the origins of the notion of partial conservativity; but the first result concerning partial conservativity is [Kreisel 62] showing that

$\neg\text{Con}^*(PA)$ is Π_1 -conservative over PA . We shall give detailed references to later works below (point (4)).

Sequential theories were introduced by [Pudlák 85]; a similar notion was investigated by [Vaught 67]. Friedman also had a similar notion but did not present an exact definition, cf. [Smoryński 85 – Friedman's research].

ACA_0 is one of various interesting subsystems of second order arithmetic; we refer to Simpson's forthcoming book on this topic [Simpson 86–90]. The model-theoretic proof of conservativity of ACA_0 over PA (1.16) is a folklore; the proof of finite axiomatizability is a variant of Gödel's proof of the fact that GB proves the schema of comprehension [Gödel 40]. See also below.

The notions of a binumeration and numeration go back to [Gödel 31] and are studied in detail in [Feferman 60]. Gödel introduced also the notion of an ω -consistent theory; this is what we called a sound theory (1.21). 1-consistent theories were introduced in [Feferman-Kreisel-Orey 62]. Rosser invented witness comparison [Rosser 36]; theorem 1.24 appears in [Tarski, Mostowski, Robinson 53].

(2) Gödel was first to construct a particular self-referential formula [Gödel 31 – Monatsh.]. The non-parametric version of the fixed-point theorems 2.1 occurs in [Carnap 34]; Gödel acknowledges Carnap's priority in [Gödel 34 – Princeton]. A parametric version occurs first in [Montague 62], where also existence of a self-referential pairs is proved (2.6). For a historical survey containing self-reference see [Smoryński 81 – fifty]. Corollary 2.3 is the celebrated result on the undefinability of truth, due to Tarski [Tarski 33, 35]. Tarski refers to [Gödel 31 – Monatsh.] for the method of diagonalization. Results of subsection (e) are classical: Gödel's main paper [Gödel 31 – Monatsh.] contains his first incompleteness theorem (as well as the statement of his second incompleteness theorem) for a system related to Whitehead-Russell's *Principia Mathematica* (1913); this paper was preceded by a short announcement [Gödel 30 – Anzeiger] and complemented by a short paper [Gödel 32 – Ergebnisse]; the last paper formulates Gödel's theorems for first order systems like our PA . Rosser's improvement (replacing ω -consistency by mere consistency) is in [Rosser 36]. Flexible formulas were invented by [Mostowski 61]; our proof follows [Kripke 62]. Gödel formulated his second incompleteness theorem for a particular system and a particular proof predicate and did not need to formulate explicit provability conditions. Provability conditions are formulated in [Hilbert-Bernays 39]. The fact that PA is not finitely axiomatizable (2.4) is due to [Ryll-Nardzewski 52]. For Löb's theorem and its motivation see [Löb 54], [Henkin 52]. A paper showing that PA cannot be axiomatized by formulas of limited arithmetical complexity is [Rabin 61]; this is a predecessor of 2.27. For Craig's theorem see [Craig 72].

Essential reflexivity of PA (2.35) was proved in [Feferman 60]. Our 2.37 is a slight improvement of a theorem of [Feferman 60]. Theorems characterizing interpretability in pure extensions of PA (2.34–2.40) are based on [Orey 59, 61], the equivalence 2.39 (i) $\Leftrightarrow$ (ii) being often called Orey's compactness theorem (cf. [Hájek 72], [Guaspari 76]).

The method of shortening definable cuts was discovered by Solovay; but he has never published the result. See [Hájek 81 – Int. II] and [Paris-Dimitracopoulos 83]; the latter paper proves that there are definable cuts having no shortening to a cut closed under exponentiation (see also [Dimitracopoulos 80]).

The rest of Sect. 3 (subsections (b)–(d)) is an elaboration of (parts of) [Pudlák 85]; only Theorem 3.20 (2) is due to Hájek. The idea of simultaneous use of two provability predicates (3.6), essential for 3.9, is due to Mycielski (unpublished).

(4) Section 4 presents results on interpretability and partial conservativity for theories containing IS_1 ; many of them were first obtained for PA . We shall refer to the original results even if they are less general than the theorem presented. An extended abstract of this section appeared as [Hájek 87 – CMUC].

4.5 (1) [Feferman 60] for $T \supseteq PA$; [Švejdar 78]. (2) [Kreisel 68] for $T \supseteq PA$; [Smoryński 80]. (3) [Kreisel 68] for $T \supseteq PA$; Švejdar (unpublished).

4.25 is due to Guaspari and Solovay (for $T \supseteq PA$): see [Guaspari 67] (their examples are more complicated than ours).

4.26 (1) is classical (cf. [Rogers 67]); (2) [Hájek 87]; (3) for $T = ZF$, $\Gamma = \Pi_1$ [Hájek 71], generalized for $T \supseteq PA$ by [Lindström 84]. (4) implicit in [Lindström 84] for $T \supseteq PA$.

4.27 For $T = ZF$, $\Gamma = \Pi_1$ Solovay (unpublished) for PA , $\Gamma = \Pi_n$ and *Consu* [Hájek 79]; for PA , $\Gamma = \Sigma_n$ and *Consu* [Quinsey 81]; in full generality for $T \supseteq PA$ [Lindström 84].

Fixed point theorem 4.29 [Shepherdson 60], [Smoryński 81 – Fifty]. Theorem 4.35 (Σ_1 -numerability) was first obtained by [Shepherdson 60].

4.55 The first example of a Σ_2 formula φ such that $(ZF + \varphi)$ is interpretable in ZF but $(GB + \varphi)$ is not in GB is in [Hájek 71] under the assumption of soundness; this assumption was removed in [Hájek, Hájková 72]. A Σ_1 formula of the desired properties was constructed by Solovay.

4.57 first appeared in [Hájek 87]; 4.50 is from [Lindström 84].

*

We shall now comment on some important topics related to the contents of Chapter III but not covered. We shall indicate only basic references, not full literature on the subject.

(a) *Reflection principles*. Roughly, a reflection principle says “if a formula is provable then it is true” (or “if all instances of a formula are provable then the formula is true”). We met an example of a provable reflection principle in I.4.34; an example of an unprovable principle is the schema

$$Pr^*_{PA}(\bar{\varphi}) \rightarrow \varphi$$

for all φ (Pr^*_{PA} being given by a Σ_1 -definition of (PA)) since its particular case (for φ being $0 = 1$) clearly implies Con^*_{PA} . On the other hand, the last schema is clearly true in N and may be used as a natural strengthening of PA . This has been studied systematically; the reader may consult [Kreisel-Lévy 68], [Schmerl 79, 80, 82]; early results on reflection principles are also discussed in detail in Smoryński's survey paper [Smoryński 77] in the Handbook of Mathematical logic.

(b) *Transfinite progressions of theories*. One can iteratively add unprovable true sentences to a sound theory, e.g. investigate the sequence PA , $PA + Con^*(PA^*)$, $PA + Con^*(PA^*) + Con^*(PA^* + Con^*(PA^*))$, ... or iteratively add some reflection principles. Using ordinal notations [Kleene 44] one can iterate transfinitely and try to characterize which part of the truth $(Th(N))$ can be grasped. See [Feferman 62], [Feferman-Spector 62], [Fenstad 68], [McAloon 80, 82].

Arithmetic and lattices: Lindström and Švejdar independently introduced and studied the lattice of interpretability types over theories T containing some arithmetic ([Lindström 79 – Aalborg, 84 – Notre Dame], [Švejdar 78]). Put $\varphi \preceq \psi$ if $(T + \varphi)$ is interpretable in $(T + \psi)$; this is a quasiorder and the corresponding factors are interpretability types; they form a distributive lattice. Many deep results about this lattice are known, in particular for $T = PA$.

[Hájková 71] and [Palúch 79] investigated lattices of consistency statements: e.g. take PA and investigate all Σ_1 binumerations of PA in PA . For each such binumeration α let Con^*_α be the corresponding consistency statement. Put $\alpha \leq \beta$ if $PA \vdash Con^*_\alpha \rightarrow Con^*_\beta$. This again gives an extremely rich lattice.

Investigation of both lattices uses extremely tricky self-referential constructions. Note that [Mycielski, Pudlák and Stern 90] investigate a lattice of interpretability types of all first order theories (not necessarily Δ_1 axiomatizable and not necessarily containing arithmetic); their lattice differs drastically from Lindström-Švejdar's lattice.

(d) *Faithful interpretations*. An interpretation $*$ of T in S is Γ -faithful if for each $\varphi \in \Gamma$, $T \vdash \varphi$ is equivalent to $S \vdash \varphi^*$. See [Feferman-Kreisel-Orey 62], [Guaspari 79] (model theoretic characterization), [Lindström 84].

(e) *Arithmetic and modal logic*. This is an extremely rich and fruitful domain: one considers formulas of the modal propositional calculus (i.e. for each formula A , we have

a formula $\Box A$ -necessarily A) and investigates their arithmetical interpretations. Such an interpretation $*$ assigns a sentence of arithmetic to each propositional variable, commutes with connectives and interprets $\Box$ as provability: $(\Box A)^*$ is $Pr^*(A^*)$ (for a fixed provability predicate for a fixed arithmetic T). There is a natural axiomatic system L (or G) satisfying (for reasonable T and Pr^*) arithmetical completeness: $G \vdash A$ iff for each arithmetical interpretation $*$, $T \vdash A^*$. A pioneering paper is [Solovay 76]; an extensive monograph is [Smoryński 85 – Self].

Further important names (see bibliography): Boolos, Sambin, de Jongh, Magari, Montagna, Artemov and others. One can introduce a further modality $\triangleright$ of interpretability and investigate modal interpretability logics and logics of partial conservativity (Švejdar, Visser, de Jongh, Veltman, Berarducci, Šavrukov, Hájek – Montagna).

(f) In this book we pay very little attention to advanced methods of *proof theory*; in particular, cut elimination is used only in Chap. V, Sect. 5. Proof theory is an extremely large domain and we shall not try to sketch its aims in a few words; instead, we refer to basic monographs: [Schütte 60], [Takeuti 71], [Pohlers 89]. See also [Schwichtenberg 77]. An application of proof theory to fragments of arithmetic is [Sieg 85]; we obtain some of the results presented by him using model theoretical rather than proof theoretical methods (in Chap. IV.)

Chapter IV

(1) The construction of a non-standard model of PA using definable ultrapower goes back to [Skolem 33] (Th. 1.7). For Łoś's lemma 1.9 see [Łoś 55]. Another early paper on models of arithmetic is [Müller 61].

Lemma 1.20 (which is the only place in this chapter explicitly using Matiyasevič's theorem) is due to [Gaifman 72] (for PA). Lemma 1.22 is due to Kirby and Theorem 1.24 elaborates a theorem of [Gaifman 72], cf. 1.25 (2). Theorem 1.29 (1) is due to [Paris-Kirby 77], (2)–(3) to Hájek, see [Hájek-Paris 86].

The method of proof of Theorem 1.33 goes back to [Paris-Kirby 78] (cf. also [Lessan 78]), but they did not deal with $P\Sigma_k$; results in 1.33 not concerning $P\Sigma_k$ are due to [Paris and Kirby 78], for the rest see [Hájek and Paris 86]. As we mentioned elsewhere, the fact that PA is not finitely axiomatizable (Corol. 1.34 (3)) was first proved by [Ryll-Nardzewski 52].

We are unable to credit an author for 1.51 and 1.52; but both theorems follow rather easily from the preceding.

Theorem 1.53 is due to [Paris-Kirby 78]. Theorem 1.59 is the famous Paris-Friedman conservation theorem; Friedman has never published his proof, Paris's proof is in [Paris 80] and in [Paris 81]. We present Kaye's proof (1.60, not published as a paper, see [Kaye 91 – Models]) and a modification of a Paris's proof (1.61) suitable for formalization in Sect. 4.

(2) We commented on the arithmetic with a top above (remark to Chap. II Sect. 2 (e)), Theorem 2.2 appears in [Cegielsky-McAloon-Wilmers 82] and is attributed to Paris. Subsections (b) and (c) present and elaborate part of [Paris 80, 81]. (See more on cuts and standard systems below.) Theorem 2.40 is due to [Ressayre 83, 86]; note that an analogous result for models of PA was first proved by [Friedman 73].

(3) Section 3 is central to Chap. IV: we present results of [Paris 80, 81]. In particular, Paris invented the notion of an indicator of a family of cuts. The term "envelope" (3.3) is borrowed from [McAloon 86]. Paris's main theorems are 3.5 and 3.7; the corollary 3.6 (characterizing $I\Sigma_1$ provably recursive functions as primitive recursive) is due independently to [Minc 76], [Takeuti 75] and Parsons (unpublished). A related paper is [Ratajczyk 88].

Section 4 presents results on model theory formalized in fragments; the result 4.8 (provability of the Paris-Friedman conservation theorem in $I\Sigma_1$) is due to Clote and

Hájek independently, cf. [Clote, Hájek, Paris 90]. The last paper contains a stronger result obtained by proof-theoretical methods.

*

Chapter IV presents only some important selected topics of model theory and its fragments. For an overview of the development of model theory of arithmetic the reader may consult [Smoryński 84 – LC82] and, with special emphasis on the work of Friedman, [Smoryński 85 – Friedman]. [Kaye 91 – Models] is a monograph devoted to models of arithmetic. We mentioned our main omission in the introduction to Chap. IV; now let us offer some references.

(a) *Standard systems.* See 2.13; for $I = N$, $SS_N(M)$ (or just $SS(M)$) is a set of subsets of N called the standard system of M (also: the set of reals of M). It is easy to show that if $M \models PA$ then $SS(M)$ is just the set of all $X \cap N$, where X is a parametrically definable subset of M . Similarly introduce $SS_0(M) = \{X \cap N \mid X \text{ nonparametrically definable in } M\}$. [Scott 62] was the first to investigate $SS_0(M)$ for $M \models PA$ ($SS_0(M)$ is the set of all subsets of N binumerable in the complete theory $Th(M)$) and characterized algebras of subsets of N obtainable as $SS_0(M)$ (often called Scott algebras). [Friedman 73] introduced $SS(M)$; it turns out that all $SS(M)$ coincide with all Scott algebras (for a particular M , $SS(M)$ and $SS_0(M)$ may or may not differ, cf. e.g. [Hájek 81 – alg]). Friedman obtained beautiful theorems on the embeddability of models in dependence on their standard systems. [Kaye 91 – Models] contains a detailed exposition.

(b) *Recursively saturated models.* A very fruitful domain. A type α in M consists of a set $\tau(x, y)$ of formulas $\varphi(x, y)$ and a parameter $b \in M$ interpreting y . The type is Δ_1 iff $\tau(x, y)$ is a Δ_1 set of formulas. The type α is *finitely satisfiable* in M if for each finite subset $\delta(x, y)$ of $\tau(x, y)$,

$$M \models (\exists a) \wedge \delta(a, b).$$

The type α is *satisfiable* in M if there is an $a \in M$ such that $M \models \varphi(a, b)$ for each $\varphi \in \tau$. M is *recursively saturated* if each finitely satisfiable Δ_1 type is satisfiable. Each model M has a recursively saturated elementary extension. For an older overview of the theory of recursively saturated models of arithmetic see [Smoryński 81 – rec-sat]; see also [Kaye 91 – Models].

(c) *Cuts in models of arithmetic.* This is another very fruitful domain; we have presented only a fraction of known results. The pioneering paper is [Kirby-Paris 77] (and Kirby's dissertation [Kirby 77]). There are lots of papers on various sorts of cuts, indicators etc.; for an older survey see e.g. [Pillay 81 – cuts] further, see e.g. [McAloon 78 – (LC77)], [Kotlarski 81, 83, 84].

(d) Note also a very interesting topic – systems of PA with parameter-free induction, i.e. $I\varphi$ for φ having just one free variable. It is customary for a fragment T of arithmetic, to denote, by T^- the corresponding subtheory having the same axiom schema(s) as T but without parameters. It is easily seen that PA^- is equivalent to PA and $I\Sigma_0^-$ is equivalent to $I\Sigma_0$; but for $n \geq 1$, $I\Sigma_n^-$ is weaker than $I\Sigma_n$ and different from III_n^- . The basic paper is [Kaye, Paris, Dimitracopoulos 88] (and Kaye's dissertation [Kaye 87]); see also [Adamowicz 88].

The above is only a selection of major topics not covered; one could continue mentioning the theory of satisfaction classes, uncountable models, indiscernibles and other developments.

Chapter V

Bounded Arithmetic was introduced by Parikh [Parikh 71]. He introduced the system which is nowadays denoted by IS_0 or ID_0 . Some weak fragments had been considered before. Shepherdson [Shepherdson 64, 65, 67] considered I_{open} . Goodstein [Goodstein 54] and Cleave and Rose [Cleave-Rose 67] studied so-called E^n -arithmetics. These systems are equational systems corresponding to Grzegorczyk classes E^n , [Grzegorczyk 53]. Another equational system was introduced by Cook [Cook 75]. His system PV is closely related to S_2^1 (this relationship is described in [Buss 86 – Bounded Arith.]). After Parikh, most of the research on IS_0 and $IS_0 + \Omega_1$ was done by Paris and Wilkie. The most important papers are [Paris-Wilkie 81 – Δ_0 Sets] and [Wilkie-Paris 87]. Let us note that [Wilkie-Paris 87] was published several years after the results of the paper were obtained. Later Buss entered this area with his book [Buss 86 – Bounded Arith.]. Several other mathematicians contributed to this field and the research is going on.

(1) The result on I_{open} is due to Shepherdson [Shepherdson 64]. For further independence results on I_{open} see [Macintyre-Marker 89], [Adamowicz 86]. Theorems 1.1 and 1.2 are due to McAloon [McAloon 82] (we present Kaye's proof of 1.1). Paris improved Theorem 1.2 to cover models of IE_1 [Paris 84]. Theorem 1.4 (only for IS_0) was proved in [Parikh 71]. The hierarchy of theories $IS_0 + \Omega_i$, $IS_0 + Exp$ and $IS_0 + Superexp$ appears in the papers of Paris and Wilkie. A system equivalent to $IS_0 + Exp$ is mentioned also in [Friedman 80]. The theory IE_1 and related systems were investigated in [Wilmers 85] and [Kaye 92 – Open]. *PIND* and *LIND* axioms were introduced in [Buss 86 – Bounded Arith.].

We mention only very briefly $\Sigma_0 PHP$, a very interesting subject with several nice results and open problems. Woods showed that $\Sigma_0 PHP$ proves Bertrand's Postulate, hence that there are infinitely many primes. (Note that by Parikh's theorem any proof of the infinitude of primes in Bounded Arithmetic must give a piece of information about the distribution of primes.) Wilkie proved a weaker version of $\Sigma_0 PHP$ in $IS_0 + \Omega_1$ which is sufficient for Woods' proof. The results are presented in [Paris-Wilkie-Woods 88]. A weak form of independence of PHP was shown by Ajtai [Ajtai 83]. He proved that if we extend IS_0 with a new relation symbol $R(x, y)$ (and do not add any special axioms about R except for the induction for the new formulae), then $PHP(R(x, y))$ is not provable in such a theory (which is denoted by $IS_0(R)$). He proved a similar result for the parity principle which says that an interval $[0, 2n + 1)$ cannot be partitioned into two-element blocks [Ajtai 90].

(2) Computational complexity is a very broad subject. In Sect. 2 we have mentioned only a few results. There are several books about this subject; we recommend the following ones [Aho-Hopcroft-Ulman 74], [Garey-Johnson 79], [Savage 76], [Wagner-Wechsung 86], [Balcazár-Díaz-Gabarró 88, 90]. The very recent Handbook of Theoretical Computer Science [van Leeuwen 90] is also a very good source. We have quite neglected an important part of complexity theory which is the complexity of boolean function; the best reference is [Wegener 87].

Here we shall only state the authors of the theorems of Sect. 2. Theorems 2.2 (a) and (b) are due to Hartmanis and Stearns [Hartmanis-Stearns 65] and Hartmanis, Lewis and Stearns [Hartmanis-Lewis-Stearns 65], respectively. Theorem 2.4 is due to Hopcroft, Paul and Valiant [Hopcroft-Paul-Valiant 75]. Theorem 2.5 was proved by Savitch [Savitch 70]. Theorem 2.6 was proved independently by Immerman and Szelepczényi [Immerman 88] and [Szelepczényi 87]. Theorem 2.6 is due to Cook [Cook 71]; in this famous paper he formulated the $P \stackrel{?}{=} NP$ problem. Let us note in passing that already in 1956 Gödel had asked a question related to the $P \stackrel{?}{=} NP$ problem in a letter to von Neumann. He asked whether one can decide the provability of a formula by a proof of length n in time linear or quadratic in n . Now we know that this problem is NP -complete. The next influential paper was [Karp 72]; it contains also the proofs of the NP -completeness of CLIQUE and HAMILTONIAN GRAPHS. The NP -completeness of the solvable equations $ax^2 + by = c$ was proved by Manders and Adleman, see [Manders 80] for a survey. Theorem 2.8 is due

to Ladner [Ladner 75]. Theorem 2.9 is due to Baker, Gill and Solovay [Baker-Gill-Solovay 75]. The Polynomial Time Hierarchy was introduced in [Stockmeyer 76]. The Linear Time Hierarchy was studied in [Wrathall 78]. The reference for Nepomnjaščij's Theorem 2.14 is [Nepomnjaščij 70]. Wrathall proved that $LinH$ is equal to the class of *rudimentary sets* introduced in [Smullyan 61]. It is not difficult to show that rudimentary sets are just Δ_0 definable sets (using the natural coding of sequences). Thus Theorem 2.16 follows from her result. Theorems 2.17 (i) and (ii) were proved in [Hartmanis-Lewis-Stearns 65] and [Hartmanis-Stearns 65] respectively. Theorem 2.18 is due to Žák [Žák 83]. Theorem 2.19 was possibly never explicitly stated, but it is a typical application of Žák's method. Theorem 2.21 is due to Pudlák. Further results on time hierarchies can be found in [Paris-Wilkie 81].

An interesting area, which we do not cover, is *counting problems*. A typical question is the following: do Σ_0 definable sets have Σ_0 definable counting functions? ($F(x)$ is a counting function for A , if $F(x)$ is the number of elements of A smaller than x .) This is closely related to the questions about *PHP* in $I\Sigma_0$. If Σ_0 sets had Σ_0 counting functions and their properties were provable in $I\Sigma_0$, then $I\Sigma_0$ would prove Σ_0 *PHP*. The relation of counting to Bounded Arithmetic and approximations of counting functions were studies in [Paris-Wilkie 85 and 87]. Recently Toda proved a result from which it follows that the sets in *PH* do not have counting functions in *PH*, provided that *PH* does not collapse [Toda 89]. Hence, assuming that *PH* does not collapse, the answer to the above question is also negative.

(3) Bennett was the first to show a Σ_0 formula for the relation $z = x^y$ [Bennett 62]. Paris found another such formula and Dimitracopoulos [Dimitracopoulos 80] verified that the inductive clauses ((c.1) and (c.2)) are provable in $I\Sigma_0$ for Paris's formula. A different Σ_0 definition of exponentiation is in [Pudlák 83 – A definition]. In Sect. 3 we use the idea of [Nelson 86]: first to build a weak form of coding of sequences (based on binary expansions) and then to define exponentiation. However Nelson does not work in $I\Sigma_0$, he works in theories interpretable in Q . (Also he uses base four expansion of a single number instead of the binary expansions of two numbers.) A formalization of syntax in $I\Sigma_0$ is considered here for the first time, though the ideas on which it is based have been around for some time. A formalization of syntax in $I\Sigma_0 + \Omega_1$ was made in [Wilkie-Paris 87]. Theorem 3.37 is new. The corollary that context-free languages belong to Σ_0^N follows also from the result of [Wrathall 78] that context-free languages are rudimentary.

(4) The definition of the theories S_2 , T_2 and their fragments S_2^i , T_2^i is due to Buss [Buss 86 – Bounded Arith.]. The theorems about the relationship between different axioms of induction and related principles are proved in that book or in [Buss 90 – Axiomatizations]; (Theorems 4.5, 4.7, 4.8, 4.10, 4.13). Our model-theoretical approach to witnessing theorems is inspired by Wilkie's proof of Buss's theorem [Wilkie 85]. Theorem 4.29 was proved by Krajíček and Takeuti, the proof here is due to Pudlák. Theorem 4.32 first appeared in [Buss 90 – Axiomatizations], it is a strengthening of Buss's theorem from [Buss 86, Bounded Arith.]. We have adapted Wilkie's proof to this strengthening. The result about models of fragments of S_2 which we have extracted from his proof (Theorem 4.31) might be of an independent interest. Theorem 4.38 is due to Krajíček, Pudlák and Takeuti [Krajíček, Pudlák and Takeuti 91]. See also [Krajíček 91 – Fragments].

A proof that S_2 is not finitely axiomatizable would strongly support the conjecture that the Polynomial Hierarchy is proper. However, proving the former statement seems to be also very hard. A possible way of proving it is proposed in [Krajíček, Pudlák 89 – Quantified]. Quantified propositional proof systems are related to the fragments S_2^i in the same way as the Extended Frege system is related to PV (hence to S_2^1) in [Cook 75] (this was done independently also by Dowd [Dowd 85]). To separate the fragments S_2^i , we need to prove a speed-up theorem about the length of proofs for the hierarchy of the quantified propositional calculi.

Versions of Buss's theorem for different fragments and different complexity function classes were proved in [Clote-Takeuti 86; 92].

The spectrum of the fragments of Bounded Arithmetic is much richer, more and more theories are being defined. A lot of research is being done on second order systems of Bounded Arithmetic. Such systems were introduced in [Buss 86 – Bounded Arith.]; for more recent results see [Takeuti 90 – S_2^1 ; 91 – A second], [Clote-Takeuti 92], [Krajčček 90 – Exponentiation].

(5) The truth definitions for Σ_0 were first considered in [Lesan 78] and [Paris-Dimitracopoulos 82]. The finite axiomatizability of $IS_0 + Exp$ (Theorem 5.6) was shown in [Gaifman-Dimitracopoulos 82]. Theorem 5.7 is due to Wilkie (unpublished). For related results on interpretability see [Szemielew-Tarski 52], [Nelson 86], (Nelson considers only local interpretations), and references to Chap. III. Theorem 5.12 is in [Wilkie-Paris 87]. We formalize some classical results of proof theory, the best reference is [Takeuti 80]. The important concept of restricted provability and restricted consistency is due to Paris and Wilkie [Wilkie-Paris 87]. Theorem 5.26 is due to Wilkie [Wilkie 86 – On sentences]. His original proof of this theorem is different, he uses a model-theoretical construction. Theorem 5.27, in a slightly stronger form, is in [Wilkie-Paris 87]. Theorem 5.28 is due to Pudlák [Pudlák 85]. Corollary 5.29 and Theorem 5.31 were proved in [Wilkie-Paris 87] (we give different proofs). Corollary 5.32 is from [Wilkie 86 – On sentences]. Theorem 5.33 and Corollary 5.34 are strengthenings of results of [Wilkie-Paris 87]. The last theorem is new.

There are more results about incompleteness and truth definitions in Bounded Arithmetic. Their main motivation is the problem of finite axiomatizability of Bounded Arithmetic. We cannot use ordinary consistency statements to separate fragments of Bounded Arithmetic, since even $IS_0 + Exp$ does not prove Con_Q^0 . In [Buss 86 – Bounded Arith.] Buss proposed to use various kinds of *bounded consistency*, which are consistencies with respect to proofs containing only bounded formulae. It has turned out that a similar incompleteness extends also to such consistency statements. For instance S_2 does not prove bounded consistency of S_2^1 [Pudlák 90 – A note], [Takeuti 88 – Some relations]. For further results see [Takeuti 88 – Bounded], [Krajčček-Takeuti to appear].

Bibliography

This bibliography grew out of our working database of papers related to our research in foundations of arithmetic. Several papers have been added when working on the book, in particular we added references to some papers and books on complexity theory. The choice of items has been considerably biased by our interests and the sources available to us. Still we believe that it can be a valuable source for mathematicians working in this field.

Our sources have been the following: First of all, we used the excellent *Omega-bibliography of mathematical logic*. This is the main source until 1984. It was impossible to copy any section from the bibliography as a whole; each item has been selected (for criteria see below). In particular, section F30 in the Omega bibliography contains many papers not included here. Our further sources were: A bibliography by Smoryński, circulated some years ago, our own works and their lists of references, most important journals and proceeding volumes of relevant conferences, our collections of preprints and reprints and some few information retrieval sessions with mathematical databases. Finally, some colleagues were sent listings of their papers contained in the bibliography and asked to send completions.

Our criteria for inclusion of a work into the bibliography have been, unfortunately, rather vague: we included papers about which we were sure or at least suspected that they were somehow relevant for topics elaborated in the book and for our future research. Since the book is devoted to the metamathematics of first-order arithmetic with special emphasis on fragments of Peano arithmetic, including weak fragments and, on the other hand, to interpretability, partial conservativity and some parts of model theory of fragments, these topics are emphasized also in the bibliography. Little attention is paid to topics as advanced proof theory or second order systems. The fact that some particular paper is not included does not mean that we hold it for irrelevant: the reason may be that the paper has been unknown to us or at least we have not known its content.

- ACKERMANN W.: Zur Widerspruchsfreiheit der Zahlentheorie, *Math. Ann.* Vol. 117, 1940 pp. 162-194
- ACZEL P.: Two notes on the Paris independence result. In: *Model Theory and Arithmetic* (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 21-31
- ADAMOWICZ Z.: Open inductions and the true theory of rationals, Preprint 289, Polish Acad. Sci., 1984
- ADAMOWICZ Z.: Axiomatization of the forcing relation with an application to PA, *Fund. Math.* Vol. 120, 1984 pp. 167-186
- ADAMOWICZ Z.: Algebraic approach to $\exists_1$ -induction. In: *Proceedings of the third Easter conference on model theory* (Gross Koris, 1985), Humboldt Univ. Berlin, 1985 pp. 5-15
- ADAMOWICZ Z.: Open induction (abstract), *Journ. Symb. Log.* Vol. 50, 1985 pp. 260
- ADAMOWICZ Z.: Some results on open and diophantine induction. In: *Logic Colloquium '84*, North Holland, 1986 pp. 1-20
- ADAMOWICZ Z.: Open induction and true theory of rationals, *Journ. Symb. Log.* Vol. 52, 1987 pp. 793-801
- ADAMOWICZ Z.: End-extending models of $IA_0 + exp + B\Sigma_1$, preprint 431, Inst. Math. Polish Acad. Sci., 1988, p. 38
- ADAMOWICZ Z.: A recursion theoretic characterization of instances of $B\Sigma_n$ provable in $\Pi_{n+1}(N)$, *Fund. Math.* Vol. 129, 1988 pp. 231-236
- ADAMOWICZ Z.: Parameter free induction, the Matiyasevič Theorem and $B\Sigma_1$. In: *Logic Colloquium '86*, North Holland, 1988 pp. 1-8
- ADAMOWICZ Z., BIGORAJSKA T.: Functions provably total in $I\Sigma_1$, *Fund. Math.*, Vol. 132, 1989, pp. 189-194
- ADAMOWICZ Z., KOSSAK R.: Note on an intermediate induction scheme. In: *Seminarberichte 71*, *Proceedings of the Fourth Easter Conference on Model Theory*, Gross-Köris, 1986 pp. 1-5
- ADAMOWICZ Z., KOSSAK R.: A note on $B\Sigma_n$ and an intermediate induction schema, *Zeitschr. Math. Log. Grundle Math.* Vol. 34, 1988 pp. 261-264
- ADAMOWICZ Z., MORALES-LUNA G.: A recursive model for arithmetic with weak induction, *Journ. Symb. Log.* Vol. 50, 1985 pp. 49-54
- ADDISON J. W.: Separation principles in hierarchies of classical and effective descriptive set theory, *Fund. Math.* Vol. 46, 1958 pp. 123-135
- ADLER A.: Extensions of non-standard models of number theory, *Zeitschr. Math. Log. Grundle Math.* Vol. 15, 1969 pp. 289-290
- AHO A. V., HOPCROFT J. E., ULLMAN J. I.: *The Design and Analysis of Computer Algorithms*, Addison-Wesley, 1974
- AJTAI M.: Σ_1^1 -formulae on finite structures, *Annals Pure Appl. Logic*, Vol. 24, 1983 pp. 1-48
- AJTAI M.: The complexity of the pigeonhole principle, 29-th Symp. on Foundations of Comp. Sci., 1988 pp. 346-355
- AJTAI M.: Parity and the pigeonhole principle. In: *Feasible Mathematics*, Buss, Scott eds., Birkhäuser 1990 pp. 1-24
- ARTEMOV S. N.: Arifmetičeski polnye modal'nye teorii. (Arithmetically complete modal theories) In: *Semiotics and information science*, No. 14, Akad. Nauk SSSR, Vsesojuz. Ins. Nauč. i Tehn. Inf., 1980 pp. 115-133 (translation in Amer. Math. Soc. Transl. (2) Vol. 135, 1987 pp. 39-54)
- ARTEMOV S. N.: Priloženia modal'noj logiki v teorii dokazatelstv. (Applications of modal logic in proof theory) In: *Nekl. logiki i ich priloženia*, Moskva, Akad. Nauk SSSR, 1982
- ARTEMOV S. N.: On modal representations of extensions of Peano arithmetic, *C. R. Math. Acad. Sci. Soc. Roy. Canada* Vol. 6, 1984 pp. 129-132
- ARTEMOV S. N.: Nearifmetičnost' istinostnyh predikatnyh logik dokazatel'nosti. (Nonarithmicity of truth predicate logics of provability.) *Doklady Akad. Nauk SSSR* Vol. 284, 1985 pp. 270-271 (translation in Soviet. Math. Dokl. Vol. 32, 1985 pp. 403-405)

- ARTEMOV S. N.: O modal'nyh logikah axiomatizujuščih dokazatel'nost'. (Modal logics axiomatizing provability.), *Izv. Akad. Nauk SSSR, Ser. Math.* Vol. 49, 1985 pp. 1123–1154
- BAAZ M., PUDLÁK P.: Kreisel's conjecture for $L\exists_1$. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 30–60
- BAKER T. P., GILL J., SOLOVAY R.: Relativizations of the $P=NP$ question, *SIAM Journ. Comput.*, Vol. 4, 1975 pp. 431–442
- BALCÁZAR J. Z., DÍAZ J., GABARRÓ J.: *Structural Complexity I*, Springer-Verlag 1988, EATCS 11, 191 p.
- BALCÁZAR J. Z., DÍAZ J., GABARRÓ J.: *Structural Complexity II*, Springer-Verlag 1990, EATCS 22, 283 p.
- BARNES D. W., MONRO G. P.: A simple model for a weak system of arithmetic, *Bull. Austral. Math. Soc.* Vol. 11, 1974 pp. 321–323
- BARWISE J., SCHLIPF J. S.: On recursively saturated models of arithmetic. In: *Model Theory and Algebra; A Memorial Tribute to Abraham Robinson*, Springer-Verlag, 1975 pp. 42–55
- BELL J. L., MACHOVER M.: A course in mathematical logic. North Holland Publ. Comp. 1977, p. 599
- BELLISIMA F.: On the modal logic corresponding to diagonalizable algebra theory, *Boll. Un. Math. Ital. B(5)* Vol. 15, 1978 pp. 915–930
- BELTYUKOV A. P.: Decidability of the universal theory of natural numbers with addition and divisibility. In: *Zap. naučn. sem. Leningr. otдел. Mat. in-ta V. A. Steklova* 60, 1976 pp. 15–27
- BELTYUKOV A. P.: A computer description and a hierarchy of initial Grzegorzczk classes, (Russian), *Zap. Nauc. Sem. LOMI* Vol. 88, 1979 pp. 30–46, (Translation in *Journ. Soviet Math.* Vol. 20, 1982)
- BENCIVENGA E.: Finitary consistency of a free arithmetic, *Notre Dame Journ. Formal Log.* Vol. 25, 1984 pp. 224–226
- BENDA M.: On strong axiom of induction in set theory and arithmetic. In: *Mathematical Logic in Latin America*, Amsterdam, North-Holland, 1980, pp. 43–56
- BENDA M.: On Harrington's partition relation, *Journ. Comb. Theory, Ser. A* Vol. 28, 1980 pp. 338–350
- BENDOVÁ K., HÁJEK P.: A logical analysis of the truth-reaction paradox, *Comm. Math. Univ. Carol.* Vol. 23, 1982 pp. 667–688
- BENNET CH.: On some orderings of extensions of arithmetic. Doctoral dissertation, University of Goteborg (Sweden)-Department of Phil, 1986 pp. 100 pages
- BENNET CH.: Lindenbaum algebras and partial conservativity, *Proc. Amer. Math. Soc.* Vol. 97, 1986 pp. 323–327
- BENNETT J. H.: Ph. D. disertation, Princeton Univ. 1962
- BERARDUCCI A.: The interpretability logic of PA, *Journ. Symb. Log.*, Vol. 55, pp. 1059–1089
- BERARDUCCI A., INTRIGILA B.: Combinatorial principles in elementary number theory, 1990 (preprint)
- BERGSTRA J. A., TUCKER J. V.: Hoare's logic and Peano's arithmetic, *Theoret. Comput. Sci.* Vol. 22, 1983 pp. 265–284
- BERLINE C.: Idéaux des anneaux de Peano (d'après Cherlin). In: *Model Theory and Arithmetic*, (Lect. Notes Math. 890), Springer-Verlag, 1982 p. 32
- BERLINE C., MCALOON K., RESSAYRE J. P. (EDS): Model theory and arithmetic. Comptes rendus d'une action thématique programmée du C. N. R. S. sur la theorie des modèles et l'arithmetique. In: (Lect. Notes Math. 890), Springer-Verlag, 1981, p. 306
- BERNARDI C.: The fixed point theorem for diagonalizable algebras, *Studia Logica* Vol. 34, 1975 pp. 239–251

- BERNARDI C.: On the equational class of diagonalizable algebras, *Studia Logica* Vol. 34, 1975 pp. 322–331
- BERNARDI C.: The uniqueness of the fixed-point in every diagonalizable algebra (The algebraization of the theories which express Theor VIII.), *Studia Logica* Vol. 35, 1976 pp. 336–343
- BERNARDI C.: On the relation provable equivalence and on partitions in effectively inseparable sets, *Studia Logica* Vol. 40, 1981 pp. 29–37
- BERNARDI C.: A shorter proof of a recent result by R. Di Paola, *Notre Dame Journ. Formal Log.* Vol. 25, 1984 pp. 390–393
- BERNARDI C., D'AQUINO P.: Topological duality for diagonalizable algebras, *Notre Dame Journ. Formal Log.* Vol. 29, 1988 pp. 345–364
- BERNARDI C., MONTAGNA F.: Equivalence relations induced by extensional formulae: classification by means of a new fixed point property, *Fund. Math.* Vol. 124, 1984 pp. 221–233
- BERNAYS P.: Über das Induktionsschema in der rekursiven Zahlentheorie. In: *Kontrolliertes Denken. Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften*. Festschrift für Wilhelm Britzelmayr, Verlag Karl Alber, Herder Freiburg, 1951 pp. 10–17
- BERNAYS P.: The original Gentzen consistency proof for number theory. In: *Intuitionism and Proof Theory*, North Holland: Amsterdam, 1970 pp. 409–417
- BEZBORUAH A., SHEPHERDSON J.: Gödel's second incompleteness theorem for Q, *Journ. Symb. Log.* Vol. 41, 1976 pp. 503–512
- BLASS A.: The intersection of nonstandard models of arithmetic, *Journ. Symb. Log.* Vol. 37, 1972 pp. 103–106
- BLASS A.: On certain types and models for arithmetic, *Journ. Symb. Log.* Vol. 39, 1974 pp. 151–162
- BLASS A.: End extensions, conservative extensions, and the Rudin-Frolik ordering, *Trans. Amer. Math. Soc.*, 1977 No. 225 pp. 325–340
- BLASS A.: Amalgamation of nonstandard models of arithmetic, *Journ. Symb. Log.* Vol. 42, 1977 pp. 372–386
- BLASS A. R.: Conservative extensions of models of arithmetic, *Arch. Math. Log.* Vol. 20, 1980 pp. 85–94
- BÖRGER E.: Computability, Complexity, Logic, North-Holland P. C. 1989, xx + 592 pp.
- BOOLOS G.: A note on Beth's theorem, *Bull. Acad. Polon. Sci.* Vol. 21, 1973 pp. 1–2
- BOOLOS G.: On Kalmar's consistency proof and a generalization of the notion of omega-consistency, *Arch. Math. Log.* Vol. 17, 1975 pp. 3–7
- BOOLOS G.: On deciding the truth of certain statements involving the notion of consistency, *Journ. Symb. Log.* Vol. 41, 1976 pp. 779–781
- BOOLOS G.: On deciding the provability of certain fixed point statements, *Journ. Symb. Log.* Vol. 42, 1977 pp. 191–193
- BOOLOS G.: Reflection principles and iterated consistency assertions, *Journ. Symb. Log.* Vol. 44, 1979 pp. 33–35
- BOOLOS G.: The unprovability of consistency. An essay in modal logic, Cambridge Univ Press: Cambridge GB, 1979, p. 184
- BOOLOS G.: Provability in arithmetic and a schema of Grzegorczyk, *Fund. Math.* Vol. 106, 1980 pp. 41–45
- BOOLOS G.: ω -consistency and the diamond, *Studia Logica* Vol. 39, 1980 pp. 237–243
- BOOLOS G.: Provability, truth and modal logic, *Journ. Phil. Logic* Vol. 9, 1980 pp. 1–7
- BOOLOS G.: On systems of modal logic with provability interpretations, *Theoria* (Lund) Vol. 46, 1980 pp. 7–18
- BOOLOS G.: Extremely undecidable sentences, *Journ. Symb. Log.* Vol. 47, 1982 p. 191
- BOOLOS G.: On the nonexistence of certain normal forms in the logic of provability, *Journ. Symb. Log.* Vol. 47, 1982 pp. 638–640
- BOOLOS G.: The logic of provability, *Amer. Math. Mon.* Vol. 91, 1984 pp. 470–480

- BOOLOS G.: 1-consistency and the diamond, *Notre Dame Journ. Formal Log.* Vol. 26, 1985 pp. 341-347
- BOOLOS G., JEFFREY R.: *Computability and Logic*, 2-nd edition, Cambridge Univ. Press, 1980, p. 262
- BOOLOS G., MCGEE V.: The degree of the set of sentences of predicate provability logic that are true under every interpretation, *Journ. Symb. Log.* Vol. 52, 1987 pp. 165-171
- BOUGHATTAS S.: *L'arithmétique ouverte et ses modèles non-standards*, These de 3^e cycle, Université Paris VII, Paris, 1987
- BOUGHATTAS S.: L'induction pour les formules ouvertes n'est pas finiment axiomatisable, *Comptes Rendus de l'Académie des Sciences de Paris*, Vol. 308 Série I, 1989 pp. 297-300
- BRANDT U.: Index sets in the arithmetical hierarchy, *Annals Pure Appl. Logic* Vol. 37, 1988 pp. 101-110
- BRITTON J. L.: Integer solutions of systems of quadratic equations. In: *Math. Proc. Cambridge Phil. Soc.* Vol. 86, 1979 pp. 385-389
- BROWN D. K., SIMPSON S. G.: Which set existence axioms are needed to prove the separable Hahn-Banach theorem?, *Annals Pure Appl. Logic* Vol. 31, 1986 pp. 123-144
- BUCHHOLZ W.: An independence results for Π_1^1 -CA + BI, *Annals Pure Appl. Logic* Vol. 33, 1987 pp. 131-155
- BUCHHOLZ W., FEFERMAN S., POHLERS W., SIEG W.: *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies.* (Lect. Notes Math. 897), Springer-Verlag, 1981, p. 383
- BUCHHOLZ W., POHLERS W.: Provable well orderings of formal theories for transfinitely iterated inductive definitions, *Journ. Symb. Log.* Vol. 43, 1978 pp. 118-125
- BUCHHOLZ W., WAINER S.: Provably computable functions and the fast growing hierarchy. In: *Logic and Combinatorics*, Amer. Math. Soc.: Arcata, 1987 pp. 179-198
- BUSS S. R.: The polynomial hierarchy and fragments of bounded arithmetic. In: 17-th Annual ACM Symposium on Theory of Computing, 1985 pp. 285-290
- BUSS S. R.: Bounded arithmetic (revised version of Ph. D. thesis), Bibliopolis, 1986
- BUSS S. R.: The polynomial hierarchy and intuitionistic bounded arithmetic. In: *Structure in Complexity Theory* (Lect. Notes Comp. Sci. 223), Springer-Verlag, 1986 pp. 77-103
- BUSS S. R.: Polynomial size proofs of the propositional pigeonhole principle, *Journ. Symb. Log.* Vol. 52, 1987 pp. 916-927
- BUSS S. R.: A conservation result concerning bounded theories and the collection axiom, *Proc. Amer. Math. Soc.* Vol. 100, 1987 pp. 109-116
- BUSS S. R.: The Boolean formula value problem is in ALOGTIME. In: 19-th Annual ACM Symp. on Theory of Computing, 1987 pp. 123-131
- BUSS S. R.: Axiomatizations and conservation results for fragments of bounded arithmetic, In: *Logic and Computation, Contemporary Mathematics 106*, Amer. Math. Soc., 1990 pp. 57-84
- BUSS S. R.: A note on bootstrapping Intuitionistic Bounded Arithmetic, *proc. Proof Theory Conference*, Leeds, England, 1990, to appear
- BUSS S. R.: The undecidability of k -provability, *Annals of Pure and Appl. Logic* 53, 1991 pp. 75-102
- BUSS S. R.: A conservation result concerning bounded theories and the collection axiom, *Proc. Amer. Math. Soc.* Vol. 100, 1987, pp. 916-927
- BUSS S. R.: Propositional consistency proofs, *Annals of Pure and Appl. Logic* 52, 1991 pp. 3-29
- BUSS S. R.: On model theory for intuitionistic bounded arithmetic with applications to independence results. In: *Feasible Mathematics*, Buss, Scott eds., Birkhäuser 1990, pp. 27-48
- BUSS S. R.: The witness function method and provably recursive fragments of Peano arithmetic. In: *Logic Methodology and Philosophy of Science '91*, to appear
- BUSS S. R. ET AL.: *Weak formal systems and connections to computational complexity*, Student written lecture notes for a course at UC Berkeley, 1988

- BUSS S. R., J. KRAJČEK J., TAKEUTI G.: Provably total functions in bounded arithmetic theories R_3^i , U_2^i and V_2^i . In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajčcek, Oxford Univ. Press 1992, pp. 116–161
- CANTINI A.: Majorizing provably recursive functions in fragments of PA, *Arch. Math. Log.* Vol. 25, 1985 pp. 21–31
- CANTOR A.: Minimal proper elementary extensions of N with respect to definable relations Dissertation, Univ. S. Carolina, 1972
- CARLSON T. J., SIMPSON S. G.: A dual form of Ramsey's theorem, preprint, Penns. Univ, 1983
- CARNAP R.: *Logische Syntax der Sprache*, Springer-Verlag, 1934
- CARSTENS H. G.: The theorem of Matijasevič is provable in Peano's arithmetic by finitely many axioms, *Logique & Anal.*, N. S. Vol. 20, 1977 pp. 116–121
- CARSTENS H. G.: Über die Kompliziertheit numerischer Modelle, Dissertation, Münster, 1972
- CARSTENS H. G.: Reducing hyperarithmetic sequences, *Fund. Math.* Vol. 89, 1975 pp. 5–11
- CARSTENS H. G.: Δ_2^0 -Mengen, *Arch. Math. Log.* Vol. 18, 1976 pp. 55–65
- CAVINESS B. F.: On canonical forms and simplification, *Journ. ACM* Vol. 17, 1970 pp. 385–396
- CEGIELSKI P.: La théorie élémentaire de la multiplication, *C. R. Acad. Sci. Paris Ser. A-B* Vol. 290, 1980 pp. A935–A938
- CEGIELSKI P.: Théorie élémentaire de la multiplication des entiers naturels. In: *Model Theory and Arithmetic*, (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 44–89
- CEGIELSKI P.: La théorie élémentaire de la divisibilité est finiment axiomatisable, *C. R. Acad. Sci. Paris Ser. I Math.* Vol. 299, 1984 pp. 367–369
- CEGIELSKI P., MCALOON K., WILMERS G.: Modeles recursivement satures de l'addition et de la multiplication des entiers naturels. In: *Logic Colloquium '80* (Studies in Logic and the Foundations of Mathematics 108), North-Holland, 1982 pp. 57–68
- CELLUCI C.: Proprieta di uniformita e 1-coerenza dell'aritmetica del primo ordine, *Matematiche* Vol. 32, 1977 pp. 35–49
- CENZER D., CLOTE P., SMITH R. L., SOARE R. I., WAINER S. S.: Members of countable Π_1^0 classes, *Annals Pure Appl. Logic* Vol. 31, 1986 pp. 145–163
- CHAITIN G. J.: Gödel's theorem and information, *Int. Journ. Theor. Phys.* Vol. 21, 1982 pp. 941–954
- CHATZIDAKIS Z.: La représentation en termes de faisceaux des modèles de la théorie élémentaire de la multiplication des entiers naturels. In: *Model theory and arithmetic* (Paris, 1979–1980) (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 90–110
- CHURCH A.: An independence question in recursive arithmetic. In: *Colloque sur les Fondements des Mathématiques, les Machines Mathématiques, et leurs Applications* (Collection de Logique Mathématique), Akad. Kiado: Budapest, 1965 pp. 21–26
- CICHON E. A.: A short proof of two recently discovered independence results using recursion theoretic methods, *Proc. Amer. Math. Soc.* Vol. 87, 1983 pp. 704–706
- CICHON E. A., WAINER S. S.: Slow growing and Grzegorzczuk hierarchies, *Journ. Symb. Log.* Vol. 48, 1983 p. 399
- CINMAN L. L.: On the complete induction axiom, *Doklady Akad. Nauk SSSR*, 1967 No. 173 pp. 273–274
- CINMAN L. L.: The role of the principle of induction in a formal arithmetical system, *Mat. Sb. (NS)*, 1968 No. 77 pp. 71–104
- CLEAVE J. P.: A hierarchy of primitive recursive functions. (Russian translation in *Problemy matematičeskoj logiki*, Mir 1970, 94–113), *Zeitschr. Math. Log. Grundle. Math.* Vol. 9, 1963 pp. 331–345
- CLEAVE J. P., ROSE H. E.: ϵ^n -Arithmetic. In: *Sets, Models a Recursion Theory*, North-Holland, 1967 pp. 297–308

- CLOTE P.: A recursion theoretic analysis of certain generalizations of Ramsey's theorem and of the Gale-Stewart theorem, Dissertation, Duke Univ, 1979
- CLOTE P.: Weak partition relations, finite games, and independence results in Peano arithmetic. In: *Model Theory of Algebra and Arithmetic* (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 92-107
- CLOTE P.: Anti-basis theorems and their relation to independence results in Peano arithmetic. In: *Model theory and arithmetic*, (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 115-133
- CLOTE P.: A note on decidable model theory. In: *Model Theory and Arithmetic*, (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 134-142
- CLOTE P.: Applications of the low basis theorem in arithmetic. In: *Recursion-theory week* (Oberwolfach, 1984), (Lect. Notes Math. 1141), Springer-Verlag, 1985 pp. 65-88
- CLOTE P.: Partition relations in arithmetic. In: *Proc. 6th Latin American Symp. on Math. Logic*, Caracas, (Lect. Notes Math. 1130), Springer-Verlag, 1985 pp. 32-68
- CLOTE P.: Ultrafilters on definable sets of arithmetic. In: *Logic Colloquium '84* (Studies in Logic and the Foundations of Mathematics 120), North-Holland, 1986 pp. 37-58
- CLOTE P.: A note on the Mac Dowell-Specker theorem, *Fund. Math.* Vol. 127, 1987 pp. 163-170
- CLOTE P.: ALOGTIME and a conjecture of S. A. Cook. In: *Logic in Computer Science*, IEEE 1990, to appear in *Annals of Mathematics and Artificial Intelligence*
- CLOTE P.: On polynomial size Frege proofs of certain combinatorial principles. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 162-184
- CLOTE P.: Bounded arithmetic and computational complexity. In: *Structure in Complexity Theory 1990*, IEEE, to appear
- CLOTE P., HÁJEK P., PARIS J.: On some formalized conservation results in arithmetic, *Arch. Math. Log.* 30, 1990, pp. 201-218
- CLOTE P., McALOON K.: Two further combinatorial theorems equivalent to the 1-consistency of Peano arithmetic, *Journ. Symb. Log.* Vol. 48, 1983 pp. 1090-1104
- CLOTE P., TAKEUTI G.: Exponential time and bounded arithmetic (extended abstract). In: *Structure in Complexity Theory* (Lect. Notes Comp. Sci. 223), Springer-Verlag, 1986 pp. 125-143
- CLOTE P., TAKEUTI G.: Bounded arithmetic for NC, ALOGTIME, L and NL, *Annals of Pure and Appl. Logic* 56, 1992 pp. 73-117
- COOK S. A.: The complexity of theorem proving procedures, 3-rd Symp. on Theory of Computing, 1971 pp. 151-158
- COOK S. A.: Feasibly constructive proofs and the propositional calculus. In: *Proc. 7-th A. C. M. STOC*, 1975 pp. 73-89
- COOK S. A., RECKHOW R. A.: The relative efficiency of propositional proof systems, *Journ. Symb. Log.* Vol. 44, 1979 pp. 36-50
- COOK S. A., URQUHART A.: Functional interpretations of feasibly constructive arithmetic. In: *Proc. ACM Symp. on Theory of Computing*, 1989, pp. 107-112
- CRAIG W.: Bases for first order theories and subtheories, *Journ. Symb. Log.* vol. 25, 1960, pp. 97-142
- CSIRMAZ L.: On definability in Peano arithmetic, *Bull. Acad. Polon. Sci.* Vol. 8, 1979 pp. 148-153
- CSIRMAZ L., PARIS J. B.: A property of 2-sorted Peano models and program verification, *Zeitschr. Math. Log. Grundle. Math.* Vol. 30, 1984 pp. 325-334
- ČUBARJAN A. A.: The lengths of derivations of formulae in extensions of formal arithmetic, *Izv. Akad. Nauk. Armjan. SSSR Ser. Mat.* 9, 1974 pp. 409-429
- ČUDA K.: An elimination of the predicate "to be a standard member" in nonstandard models of arithmetic, *Comm. Math. Univ. Carol.* Vol. 23, 1982 pp. 785-803
- ČUDNOVSKY G. V.: Diophantine predicates, *Uspekhi Matem. Nauk.* Vol. 25, 1970 No. 4 pp. 185-186

- DAGUENET M.: Un modèle non standard de l'arithmétique, *Comptes Rendus*, 1972 No. 274 pp. 685-688
- VAN DALEN D., GORDON C. E.: Independence problems in subsystems of intuitionistic arithmetic, *Indag. Math.* Vol. 33, 1971 pp. 448-456
- VAN DALEN D., MULDER H., KRABBE E. C. W., VISSER A.: Finite Kripke models of HA are locally PA, *Notre Dame Journ. Formal Log.* Vol. 27, 1986 pp. 528-532
- DAVIS M. D.: Arithmetical problems and recursively enumerable predicates, *Journ. Symb. Log.* Vol. 18, 1953 pp. 33-41
- DAVIS M. D.: One equation to rule them all, *Trans. N. Y. Acad. Sci.* Vol. 30, 1960 pp. 766-775
- DAVIS M. D.: Application of recursive function theory to number theory. In: *Recursive Function Theory*, Amer. Math. Soc., Providence, 1962, pp. 135-138
- DAVIS M. D.: Extensions and corollaries of recent work on Hilbert's tenth problem, III. *Journ. Math.* Vol. 7, 1963 pp. 246-250
- DAVIS M. D.: Introduction to Gödel's paper: On formally undecidable propositions of the Principia Mathematica and related systems I. In: *The Undecidable. Basic papers on Undecidable Propositions, Unsolvable problems and Computable functions*, Raven Press, New York, 1965 pp. 4
- DAVIS M. D.: Diophantine equations and recursively enumerable sets. In: *Automata Theory*, New York, Academic Press, 1966, pp. 146-152
- DAVIS M. D.: An explicit Diophantine definition of the exponential function, *Commun. Pure and Applied Math.* Vol. 24, 1971 pp. 137-145
- DAVIS M. D.: On the number of solutions of Diophantine equations. In: *Proc. Amer. Math. Soc.* Vol. 33, 1972 pp. 552-554
- DAVIS M. D.: Hilbert tenth problem is unsolvable, *Amer. Math. Mon.* Vol. 80, 1973 pp. 233-269
- DAVIS M. D.: Speed-up theorems and Diophantine equations. In: *Courant Computer Science Symposium 7: Computational Complexity*, New York, Algorithmics Press, 1973
- DAVIS M. D.: What is a computation? In: *Mathematics Today*, Heidelberg, Springer-Verlag, 1978
- DAVIS M. D. (ED): *The undecidable. Basic papers on undecidable propositions, unsolvable problems and computable functions*, Raven Press, New York, 1965, 440 p.
- DAVIS M. D., MATIYASEVIČ I., ROBINSON J.: Hilbert's tenth problem. Diophantine equations: positive aspects of a negative solution. In: *Mathematical Developments Arising From Hilbert's Problems*, Amer. Math. Soc. Proc. Symp. Pure Math. 28, 1976, pp. 223-378
- DAVIS M. D., PUTNAM H.: Reduction of Hilbert's tenth problem, *Journ. Symb. Log.* Vol. 23, 1958 pp. 183-187
- DAVIS M. D., PUTNAM H.: Diophantine sets over polynomial rings, III. *Journ. Math.* Vol. 7, 1963 pp. 251-256
- DAVIS M. D., PUTNAM H., ROBINSON J.: The decision problem for exponential Diophantine equations, *Ann. Math.* Vol. 74, 1961 pp. 425-436
- DAWES A. M., FLORENCE J. B.: Independent Gödel sentences and independent sets, *Journ. Symb. Log.* Vol. 40, 1975 pp. 159-166
- DAWES A. M.: First order hierarchies in general models and in models of Peano arithmetic, *Dissertation*, Toronto, 1972
- DAWES A. M.: End extensions which are models of a given theory, *Zeitschr. Math. Log. Grundle Math.* Vol. 23, 1977 pp. 463-467
- DAWSON J.: The Gödel incompleteness theorem from a length-of-proof perspective, *Amer. Math. Mon.* Vol. 86, 1979 pp. 740-748
- DEDEKIND R.: Was sind und was sollen die Zahlen, Braunschweig, F. Vieweg und Sohn, 1939, 58 pp.
- DENEFF J.: Hilbert's tenth problem for quadratic rings, *Proc. Amer. Math. Soc.* Vol. 48, 1975 pp. 214-220

- DENEJ J.: Diophantine sets over $\mathbb{Z}(T)$. In: Proc. Amer. Math. Soc. Vol. 69, 1978 pp. 148-150
- DENEJ J.: The Diophantine problem for polynomial rings and fields of rational functions, Trans. Amer. Math. Soc., 1978 No. 242 pp. 391-399
- DENEJ J.: The Diophantine problem for polynomial rings of positive characteristic. In: Logic Colloquium '78, Amsterdam, North-Holland, 1979, pp. 131-145
- DENEJ J.: Diophantine sets over algebraic number rings II, Trans. Amer. Math. Soc., 1980 No. 257 pp. 227-236
- DENEJ J., LIPSHITZ L.: Diophantine sets over some rings of algebraic integers, Journ. London Math. Soc. Vol. (2) 18, 1978 pp. 385-391
- DENIS R.: The arithmetics as theories of two orders. In: Orders: descriptions and roles (L'Arbresle, 1982), North-Holland, 1984 pp. 287-311
- DENISOV S. D.: Models of noncontradictory formulas and the Ershov hierarchy, Algebra i Logika Vol. 11, 1972 pp. 648-655
- DEREVYANKINA E. A.: Some questions on the completeness of arithmetic, Sibirski Mat. Zh. Vol. 15, 1974 pp. 299-317
- DIACONESCU R., KIRBY L. A. S.: Models of arithmetic and categories with finiteness conditions, Annals Pure Appl. Logic Vol. 35, 1987 pp. 123-148
- DICKMAN M.: Types remarquables et extensions de modèles dans l'arithmétique de Peano I, Astérisque Vol. 73, 1980 pp. 59-117
- DIMITRACOPOULOS C.: Matiyasevič theorem and fragments of arithmetic, Dissertation Manchester, 1980
- DIMITRACOPOULOS C.: A generalization of a theorem of H. Friedman, Zeitschr. Math. Log. Grundle. Math. Vol. 31, 1985 pp. 221-225
- DIMITRACOPOULOS C., GAIFMAN H.: Fragments of Peano's arithmetic and the MRDP theorem. In: Logic and Algorithmic. An International Symposium Held in Honour of Ernst Specker, Enseign. Math. Univ Genève: Genève, 1982 pp. 187-206
- DIMITRACOPOULOS C., PARIS J. B.: Truth definitions for Δ_0 formulae. In: Logic and Algorithmic. An International Symposium Held in Honour of Ernst Specker, Monogr. Enseign. Math. Univ Genève, 1982 pp. 317-329
- DIMITRACOPOULOS C., PARIS J. B.: A note on the undefinability of cuts, Journ. Symb. Log. Vol. 48, 1983 pp. 564-569
- DIMITRACOPOULOS C., PARIS J. B.: The pigeon hole principle and fragments of arithmetic, Zeitschr. Math. Log. Grundle. Math. Vol. 32, 1986 pp. 73-80
- DIMITRACOPOULOS C., PARIS J. B.: The pigeonhole principle and fragments of arithmetic. (abstract), Journ. Symb. Log. Vol. 51, 1986 pp. 485
- DIMITRACOPOULOS C., PARIS J. B.: A note on a theorem of H. Friedman, Zeitschr. Math. Log. Grundle. Math. Vol. 34, 1988 pp. 13-17
- DOWD M.: Propositional representation of arithmetic proofs (dissertation), Univ. of Toronto, 1979
- DOWD M.: Model theoretic aspects of PNP (preprint), 1985
- VAN DEN DRIES L.: Some model theory and number theory for models of weak systems of arithmetic. In: Model Theory of Algebra and Arithmetic (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 346-362
- EHRENFUCHT A.: Polynomial functions with exponentiation are well ordered, Algebra Universalis Vol. 3, 1973 pp. 261-262
- EHRENFUCHT A.: Discernible elements in models for Peano arithmetic, Journ. Symb. Log. Vol. 38, 1973 pp. 291-292
- EHRENFUCHT A., FEFERMAN S.: Representability of recursively enumerable sets in formal theories, Arch. Math. Log. Vol. 5, 1959 pp. 37-41
- EHRENFUCHT A., JENSEN D. C.: Some problems in elementary arithmetics, Fund. Math. Vol. 92, 1976 pp. 223-245
- EHRENFUCHT A., KREISEL G.: Strong models for arithmetic, Bull. Acad. Polon. Sci. Vol. 14, 1966 pp. 107-110

- EHRENFEUCHT A., MYCIELSKI J.: Abbreviating proofs by adding new axioms, *Bull. Amer. Math. Soc.* Vol. 77, 1971 pp. 366-367
- ELLENTUCK E.: A model of arithmetic, *Bull. Acad. Polon. Sci.* Vol. 22, 1974 pp. 353-355
- ENDERTON H. B.: On provable recursive functions, *Notre Dame Journ. Formal Log.* Vol. 9, 1968 pp. 86-88
- ERDOS P., MILLS G.: Some bounds for the Ramsey-Paris-Harrington numbers, *Journ. Combin. Theory Ser. A* Vol. 30, 1981 pp. 53-70
- ESENINE-VOLPINE A. S.: Le programme ultra-intuitionniste des fondements des mathématiques. In: *Infinitistic Methods, Proceedings of the Symposium on Foundations of Mathematics*, PWN: Warsaw, 1961 pp. 201-223
- FARKAS E. J., SZABO M. E.: On the programs-as-formulas interpretation of parallel programs in Peano arithmetic, *Annals Pure Appl. Logic* Vol. 37, 1988 pp. 111-127
- FEFERMAN S.: Formal consistency proofs and interpretability of theories. In: *Summaries of Talks Presented at the Summer Institute for Symbolic Logic*; 1957 Ithaca, Institute for Defense Analyses, Com. Res. Division, 1957 pp. 71-77
- FEFERMAN S.: Arithmetization of metamathematics in a general setting, *Fund. Math.* Vol. 49, 1960 pp. 35-92
- FEFERMAN S.: Transfinite recursive progressions of axiomatic theories, *Journ. Symb. Log.* Vol. 27, 1962 pp. 259-316
- FEFERMAN S.: Autonomous transfinite progressions and the extent of predicative mathematics. In: *Proceedings of 3rd International Congress for Logic, Methodology and Philosophy of Science (Stud. Logic Found. Math.)*, North Holland: Amsterdam, 1968 pp. 121-135
- FEFERMAN S.: Theories of finite type related to mathematical practice. In: *Handbook of Mathematical Logic*, Amsterdam, North-Holland, 1977, pp. 913-971
- FEFERMAN S.: Inductively presented systems and the formalization of meta-mathematics. In: *Logic Colloquium '80 (Stud. Logic Found. Math. 108)*, North Holland: Amsterdam, 1982 pp. 95-128
- FEFERMAN S.: Intensionality in mathematics, *Journ. Philosophical Logic* Vol. 14, 1985 pp. 41-56
- FEFERMAN S., KREISEL G., OREY S.: 1-consistency and faithful interpretations, *Arch. Math. Log.* Vol. 6, 1962 pp. 52-63
- FEFERMAN S., SPECTOR C.: Incompleteness along paths in progressions of theories, *Journ. Symb. Log.* Vol. 27, 1962 pp. 383-390
- FEFERMAN S., KREISEL G., OREY S.: 1-consistency and faithful interpretations, *Arch. Math. Log.* Vol. 6, 1962 No. 6 pp. 52-63
- FENSTAD J. E.: On the completeness of some transfinite recursive progressions of axiomatic theories, *Journ. Symb. Log.* Vol. 33, 1968 pp. 69-76
- FENSTAD J. E.: Non-standard models for arithmetic and analysis. In: *Proceedings of the 15th Scandinavian Congress (Lect. Notes Math. 118)*, Springer-Verlag, 1970 pp. 30-47
- FISCHER M. J., RABIN M. O.: Super-exponential complexity of Presburger arithmetic. In: *Complexity of computation*, SIAM-Amer. Math. Soc. Proceeding 7, Providence, 1974, pp. 27-44
- FISCHER P. C.: Theory of provably recursive functions. *Trans. Amer. Math. Soc.* Vol. 117, 1965 pp. 494-520
- FLAGG R. C., FRIEDMAN H. M.: Epistemic and intuitionistic formal systems, *Annals Pure Appl. Logic* Vol. 32, 1986 pp. 53-60
- FLANNAGAN, T. B.: A note on a proof of Shepherdson, *Arch. Math. Log.* Vol. 23, 1983 pp. 55-60
- FRIEDMAN H. M.: Iterated inductive definitions and Σ_2^1 -AC. In: *Intuitionism and proof theory*, Amsterdam, North-Holland, 1970, pp. 435-442
- FRIEDMAN H. M.: Countable models of set theories. In: *Cambridge Summer School in Mathematical Logic (Lect. Notes Math. 337)*, Springer-Verlag, 1973, pp. 539-573

- FRIEDMAN H. M.: Provable equality in primitive recursive arithmetic with and without induction, *Pac. Journ. Math.* Vol. 57, 1975 pp. 379-392
- FRIEDMAN H. M.: One hundred and two problems in mathematical logic, *Journ. Symb. Log.*, Vol. 40, 1975, pp. 113-129
- FRIEDMAN H. M.: Set theoretic foundations for constructive analysis, *Ann. Math.* Vol. 105, 1977 pp. 1-28
- FRIEDMAN H. M.: Classically and intuitionistically provably recursive functions. In: *Higher Set Theory* (Lect. Notes Math. 669), Springer-Verlag, 1978 pp. 21-27
- FRIEDMAN H. M.: Simpler combinatorial theorem independent of set theory, Preprint, Ohio State Univ., 1979, 6 p.
- FRIEDMAN H. M.: On the consistency, completeness, and correctness problems, Preprint, Ohio State Univ. 1979, 10 p.
- FRIEDMAN H. M.: A strong conservative extension of Peano arithmetic. In: *The Kleene Symposium* (Stud. Logic Found. Math. 101), North Holland: Amsterdam, 1980 pp. 113-122
- FRIEDMAN H. M.: Translatability and relative consistency II. Preprint, Ohio State Univ. 1980, 6 p. (see [SMORYŃSKI 82, Nonstandard models])
- FRIEDMAN H. M., MCALOON K., SIMPSON S.: A finite combinatorial principle which is equivalent to the 1-consistency of predicative analysis. In: *Patras Logic Symp*, 1982 pp. 197-230
- FRIEDMAN H. M., SHEARD M.: An axiomatic approach to self-referential truth, *Annals Pure Appl. Log.* Vol. 33, 1987 pp. 1-21
- FRIEDMAN H. M., SHEARD M.: The disjunction and existence properties for axiomatic systems of truth, *Annals Pure Appl. Log.* Vol. 40, 1988 pp. 1-10
- FRIEDMAN H. M., SIMPSON S. G., SMITH R. L.: Countable algebra and set existence axioms, *Ann. Pure Appl. Log.* Vol. 25(2), 1983 pp. 141-182
- FRIEDMAN H. M., SIMPSON S. G., SMITH R. L.: Addendum to "Countable algebra and set existence axioms", *Annals Pure Appl. Log.* Vol. 28, 1985 pp. 319-320
- FRIEDRICHS DORF U.: Einige Bemerkungen zur Peano Arithmetik, *Zeitschr. Mat. Log. Grundle. Math.* Vol. 22, 1976 pp. 431-436
- GAIFMAN H.: Uniform extension operators for models and their applications. In: *Sets, Models and Recursion Theory*, Amsterdam, North-Holland, 1967, pp. 122-155
- GAIFMAN H.: On local arithmetical functions, and their applications for constructing types of Peano's arithmetic. In: *Mathematical Logic and Foundations of Set Theory*, Amsterdam, North-Holland, 1970, pp. 105-121
- GAIFMAN H.: A note on models and submodels of arithmetic. In: *Conference in Mathematical Logic*, London '70, Springer-Verlag, 1972 pp. 128-144
- GAIFMAN H.: Models and types of Peano's arithmetic, *Arch. Math. Log.* Vol. 9, 1976 pp. 223-306
- GAIFMAN H., DIMITRACOPOULOS C.: Fragments of Peano's arithmetic and the MRDP theorem. In: *Logic and algorithmic* (Zurich, 1980) (Monograph. Enseign. Math. 30), Genève, Univ. Genève, 1982 pp. 187-206
- GANDY R. O.: Note on a paper of Kemeny's, *Math. Ann.*, 1958 No. 136 pp. 466
- GAREY M. R., JOHNSON D. S.: *Computers and Intractability: A Guide to the theory of NP-completeness*, Freeman, San Francisco, 1979, 338 pp.
- GANDY R. O.: Limitations to mathematical knowledge. In: *Logic Colloquium '80* (Studies in Logic and the Foundations of Mathematics 108), North-Holland, 1982 pp. 129-146
- GARRO I.: Independence proofs in arithmetic theories with very weak induction, *Univ Math Inst: Bonn*, 1973, 71 p.
- GARRO I.: Nonstandard models for a fragment of the arithmetic and their decision problem, *Zeitschr. Math. Log. Grundle. Math.* Vol. 33, 1987 pp. 481-483
- GEISER J. R.: A formalization of Essenin-Volpin's proof theoretical studies by means of nonstandard analysis, *Journ. Symb. Log.* Vol. 39, 1974 pp. 81-87

- GENTZEN G.: Untersuchungen über das logische Schliessen I. *Math. Zeitschr.* 39, 1935 pp. 176–210 (Russian translation in *Matematicheskaja teorija logiceskogo vyvoda*, Nauka 1967, 9–76)
- GENTZEN G.: Untersuchungen über das logische Schliessen II. *Math. Zeitschr.* Vol. 39, 1935 pp. 405–443 (Russian translation in *Matematicheskaja teorija logiceskogo vyvoda*, Nauka 1967, 9–76)
- GENTZEN G.: Die Widerspruchsfreiheit der reinen Zahlentheorie. *Math. Ann.* Vol. 112, 1936 pp. 493–565 (Russian translation in *Matematicheskaja teorija logiceskogo vyvoda*, Nauka 1967, 77–153)
- GENTZEN G.: Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften* Vol. 4, 1938 pp. 19–44 (Russian translation in *Matematicheskaja teorija logiceskogo vyvoda*, Nauka 1967, 154–190)
- GENTZEN G.: Zusammenfassung von mehreren vollständigen Induktionen zu einer einzigen, *Arch. Math. Log.* Vol. 2, 1954 pp. 1–3
- GENTZEN G.: *The collected papers of Gerhard Gentzen*, North Holland: Amsterdam, 1969, 338 p.
- GENTZEN G.: Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie, *Arch. Math. Log.* Vol. 16, 1974 pp. 97–118
- GENTZEN G.: Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik, *Arch. Math. Log.* Vol. 16, 1974 pp. 119–132
- GERMANO G.: Incompleteness theorem via weak definability of truth: a short proof, *Notre Dame Journ. Formal Log.* Vol. 14, 1973 pp. 377–380
- GINSBURG S., SPANIER E.: Semigroups, Presburger formulas and languages, *Pac. Journ. Math.* Vol. 16, 1966 pp. 285–296
- GÖDEL K.: Die Vollständigkeit der Axiome des logischen Funktionenkalküls, *Monatshefte für Math. und Phys.* vol 37, 1930, pp. 349–360.
- GÖDEL K.: Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit. In: *Math. Naturw. Klasse, Anzeiger*, 67, Akad. Wiss. Wien, 1930 pp. 214–215
- GÖDEL K.: Remarks contributed to a "Diskussion zur Grundlegung der Mathematik", *Erkenntnis* Vol. 2, 1931 pp. 147–148
- GÖDEL K.: Eine Interpretation des intuitionistischen Aussagenkalküls, *Ergebnisse eines math. Kolloquiums* Vol. 4, 1931 pp. 39–40
- GÖDEL K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, *Monatshefte Math. Phys.* Vol. 38, 1931 pp. 173–198
- GÖDEL K.: Über Vollständigkeit und Widerspruchsfreiheit, *Ergebnisse eines Mathematischen Kolloquiums* Vol. 3, 1932 pp. 12–13
- GÖDEL K.: Zu intuitionistischen Arithmetik und Zahlentheorie. In: *Ergebnisse eines Math. Kolloquiums*, 1933 pp. 34–38
- GÖDEL K.: On undecidable propositions of formal mathematical systems (mimeographed lecture notes by S. C. Kleene and J. B. Rosser). Princeton 1934 (reprinted in Davis 1965 and Gödel 1986)
- GÖDEL K.: Über die Länge von Beweisen. In: *Ergeb. eines math. Kolloq.* Vol. 7, 1936 pp. 23–24
- GÖDEL K.: The consistency of the axiom of choice and of the generalized continuum hypothesis. *Proc Nat Acad Sci USA* Vol. 24, 1938 pp. 556–557
- GÖDEL K.: Über eine bisher noch nicht benutzte Erweiterung des finiten Standpunktes, *Dialectica* Vol. 12, 1958 pp. 280–287
- GÖDEL K.: On undecidable propositions of formal mathematical systems. In: *The Undecidable*, N. Y, Raven Press, Hewlett, 1965, pp. 41–71
- GÖDEL K.: *Collected works*, vol. I (edited by S. Feferman et al.). Oxford Univ. Press New York and Oxford, 1986, xvi + 474 pp.

- GÖDEL K.: Collected works, vol. II (edited by S. Feferman et al.). Oxford Univ. Press New York and Oxford, 1990, xvi + 407 pp.
- GOLD B.: Relatively diophantine correct models of arithmetic, *Notre Dame Journ. Formal Log.* Vol. 28, 1987 pp. 291–296
- GOLDBLATT R.: Arithmetic necessity, provability and intuitionistic logic, *Theoria* Vol. 44, 1978 pp. 38–46
- GOLDFARB W. D.: On the effective ω -rule, *Zeitschr. Math. Log. Grundle. Math.* Vol. 21, 1975 pp. 409–412
- GOLDFARB W. D.: Ordinal bounds for k -consistency, *Journ. Symb. Log.* Vol. 39, 1978 pp. 693–699
- GOLDFARB W. D., SCANLON T. M.: The ω -consistency of number theory via Herbrand's theorem, *Journ. Symb. Log.* Vol. 39, 1974 pp. 678–692
- GOLDFREI D. C., MACINTYRE A., SIMMONS H.: The forcing companions of number theories, *Israel Journ. Math.* Vol. 14, 1973 pp. 318–337
- GOODMAN N. D.: Flagg realizability in arithmetic, *Journ. Symb. Log.* Vol. 51, 1986 pp. 387–392
- GOODSTEIN R. L.: On the restricted ordinal theorem, *Journ. Symb. Log.* Vol. 9, 1944 pp. 33–41
- GOODSTEIN R. L.: Transfinite ordinals in recursive number theory, *Journ. Symb. Log.* Vol. 12, 1947 pp. 123–129
- GOODSTEIN R. L.: Logic-free formalism of recursive arithmetic, *Math. Scand.* Vol. 2, 1954, pp. 247–261
- GORDEEV L.: Proof-theoretical analysis: weak systems of functions and classes, *Annals Pure Appl. Logic* Vol. 38, 1988 pp. 1–121
- GORJAČEV S. V.: On interpretability of some extensions of arithmetic, *Matematicheskie Zametki* Vol. 40, 1986 pp. 561–571
- GRATTAN-GUINNESS I.: In memorium Kurt Gödel: His 1931 correspondence with Zermelo on his incompleteness theorem, *Hist. Mathematica* Vol. 6, 1979 pp. 294–304
- GRILLIOT T. J.: Disturbing arithmetic, *Journ. Symb. Log.* Vol. 50, 1985 pp. 375–379
- GRZEGORCZYK A.: Some classes of recursive functions. In: *Rozprawy Matematyczne*, IV, Warszawa, 1953 (Russian translation in *Problemy matematicheskoy logiki*, Mir 1970, 9–49)
- GRZEGORCZYK A.: Some proofs of undecidability of arithmetic, *Fund. Math.* Vol. 43, 1956 pp. 166–177
- GRZEGORCZYK A.: An unfinetizability proof by means of restricted reduced power, *Fund. Math.* Vol. 73, 1971 pp. 37–49
- GRZEGORCZYK A.: *Zarys arytmetyki teoretycznej*. (Second edition), Państwowe Wydawnictwo Naukowe, 1983, 315 p.
- GRZEGORCZYK A., MOSTOWSKI A., RYLL-NARDZEWSKI C.: The classical and the ω -complete arithmetic, *Journ. Symb. Log.* Vol. 23, 1958 pp. 188–206
- GUASPARI D.: Partially conservative extensions of arithmetic, *Transactions Amer. Math. Soc.* Vol. 254, 1979 pp. 47–68
- GUASPARI D.: Sentences implying their own provability, *Journ. Symb. Log.* Vol. 48, 1983 pp. 777–789
- GUASPARI D., SOLOVAY R. M.: Rosser sentences, *Ann. Math. Log.* Vol. 16, 1979 pp. 81–99
- HÁJEK P.: Syntactic models of axiomatic theories, *Bull. Acad. Polon. Sci.* Vol. 13, 1965 pp. 273–278
- HÁJEK P.: Generalized interpretability in terms of models, *Čas. pěst. mat.* Vol. 91, 1966 pp. 352–357
- HÁJEK P.: On interpretability in set theories, *Comm. Math. Univ. Carol.* Vol. 12, 1971 pp. 73–79
- HÁJEK P.: On interpretability in set theories II, *Comm. Math. Univ. Carol.* Vol. 13, 1972 pp. 445–455

- HÁJEK P.: Experimental logics and Pi_3^0 theories, *Journ. Symb. Log.* Vol. 42, 1977 pp. 515–522
- HÁJEK P.: Arithmetical complexity of some problems in computer science. In: *Math. Foundations of Computer Science (Lect. Notes Comp. Sci. 53)*, Springer-Verlag, 1977 pp. 282–297
- HÁJEK P.: Arithmetical hierarchy and complexity of Computations, *Theor. Comp. Sci.* Vol. 8, 1979 pp. 227–237
- HÁJEK P.: On partially conservative extensions of arithmetic. In: *Logic Colloquium 78*, North-Holland Pub, 1979 pp. 225–234
- HÁJEK P.: Making dynamic logic first order. In: *Mathematical foundations of Computer Science*, (Lect. Notes Comp. Sci. 118), Springer-Verlag, 1981 pp. 287–295
- HÁJEK P.: On interpretability in theories containing arithmetic II, *Comm. Math. Univ. Carol.* Vol. 22, 1981 pp. 617–688
- HÁJEK P.: Completion closed algebras and models of Peano arithmetic, *Comm. Math. Univ. Carol.* Vol. 22, 1981 pp. 585–594
- HÁJEK P.: Arithmetical interpretations of dynamic logic, *Journ. Symb. Log.* Vol. 48, 1983 pp. 704–713
- HÁJEK P.: On a new notion of partial conservativity. In: *Computation and proof theory, Proc. Logic Colloq, Aachen 1983, Part II*, (Lect. Notes Math. 1104), Springer-Verlag, 1984 pp. 217–232
- HÁJEK P.: Some conservativeness results for non-standard dynamic logic. In: Demetrevics (ed.), *Proc. Conf. on Algebra, Combinatorics and Logic in Comp. Sci*, North-Holland, 1986 pp. 443–449
- HÁJEK P.: A simple dynamic logic, *Theor. Comp. Sci.* Vol. 46, 1986 pp. 239–259
- HÁJEK P.: Partial conservativity revisited, *Comm. Math. Univ. Carol.* Vol. 28, 1987 pp. 679–690
- HÁJEK P.: Partial conservativity revisited. (abstract) *Journ. Symb. Log.* Vol. 52, 1987 pp. 1066
- HÁJEK P.: Interpretability and fragments of arithmetic. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 185–196
- HÁJEK P., HÁJKOVÁ M.: On interpretability in theories containing arithmetic, *Fund. Math.* Vol. 76, 1972 pp. 131–137
- HÁJEK P., KUČERA A.: On recursion theory in $I\Sigma_1$, *Journ. Symb. Log.* Vol. 54, 1989, pp. 576–589
- HÁJEK P., MONTAGNA F.: ILM is the logic of Π_1 conservativity, *Arch. Math. Log.* Vol. 30, 1990, pp. 113–123
- HÁJEK P., MONTAGNA F., PUDLÁK P.: Abbreviating proofs using metamathematical rules. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 197–221
- HÁJEK P., PARIS J. B.: Combinatorial principles concerning approximation of functions, *Arch. Math. Log.* Vol. 26, 1986 pp. 13–28
- HÁJEK P., PUDLÁK P.: Two orderings of the class of all countable models of arithmetic. In: *Model theory of algebra and arithmetic*, (Lect. Notes Math. 843), Springer-Verlag, 1980 pp. 174–185
- HÁJKOVÁ M.: The lattice of binumerations of arithmetic, *Comm. Math. Univ. Carol.* Vol. 12, 1971 pp. 81n, 281n.
- HANDLEY W. G.: Some machine characterizations of classes close to Δ_0^N , Ph. D. dissertation, Manchester University, Manchester, 1986 pp. iv+106 pp
- HANDLEY W. G., PARIS J. B., WILKIE A. J.: Characterizing some low arithmetic classes. In: *Theory of algorithms (Pecs)*, Coll. Math. Soc. Janos Bolyai 44, 1984 pp. 353–364
- HARDY G. H., WRIGHT E. M.: *An introduction to the theory of numbers*, (3. ed.), Oxford Univ. Press, 1954

- HARNIK V.: ω_1 -like recursively saturated models of Presburger's arithmetic, *Journ. Symb. Log.* Vol. 51, 1986 pp. 421-429
- HARRINGTON L. A., PARIS J. B.: A mathematical incompleteness in Peano arithmetic. In: *Handbook of Mathematical Logic* (Stud Logicae Fundam Math 90), North Holland: Amsterdam, 1977 pp. 1133-1142
- HART J.: ε_0 -arithmetic, *Zeitschr. Math. Log. Grundle Math.* Vol. 15, 1969 pp. 237
- HARTMANIS J., HOPCROFT J. E.: Independence results in computer science, *SIGACT News* 8, 1976 No. 4 pp. 13-24
- HARTMANIS J., LEWIS P. M., STEARNS R. E.: Hierarchies of memory limited computations, *IEEE Conference Record on Switching Circuit Theory and Logical Design*, Ann. Arbor, 1965 pp. 179-190
- HARTMANIS J., STEARNS R. E. On the computational complexity of algorithms, *Trans. Amer. Math. Soc.*, Vol. 117, 1965 pp. 285-306
- HAUSCHILD K.: Über die Charakterisierbarkeit der Zahlenreihe in gewissen Nichtstandardmodellen der Arithmetik, *Zeitschr. Math. Log. Grundle Math.* Vol. 9, 1963 pp. 113-116
- HAUSCHILD K.: Nichtaxiomatisierbarkeit von Satzmengen durch Ausdrücke spezieller Gestalt, *Fund. Math.* Vol. 72, 1971 pp. 245-253
- VAN HEIJENOORT J. (ED.): From Frege to Gödel, A Source Book in Mathematical Logic, 1879-1931. Harvard Univ. Press 1967, 660 p.
- HENKIN L.: A problem concerning provability, *Journ. Symb. Log.* Vol. 17, 1952 pp. 160
- HENSON C. W., KAUFMANN M., KEISLER H. J.: The strength of nonstandard methods in arithmetic, *Journ. Symb. Log.* Vol. 49, 1984 pp. 1039-1058
- HERBRAND J.: Non-contradiction des axiomes arithmétiques, *C. R. Acad Sci, Paris* Vol. 188, 1929 pp. 303-304
- HERBRAND J.: Recherches sur la théorie de la démonstration (thesis, Paris), *Prace Towarzystwa Naukowego Warszawskiego* vol. III no. 33, 1930. (Chap. 5 see van Heijenoort 1967).
- HERBRAND J.: Sur la non-contradiction de l'arithmétique, *Journal für die Reine und Angewandte Mathematik* Vol. 166, 1931 pp. 1-8
- HERMES H.: Ein mengentheoretisches Modell des Peanoschen Axiomensystem, *Math Unterricht* Vol. 13/3, 1967 pp. 5-31
- HILBERT D., ACKERMANN W.: *Grundzüge der theoretischen Logik*. Springer-Verlag, 1928, 120 pp.
- HILBERT D., BERNAYS P.: *Grundlagen der Mathematik I*, Springer-Verlag, 1934, 471 p.
- HILBERT D., BERNAYS P.: *Grundlagen der Mathematik II*, Springer-Verlag, 1939, 561 p.
- HIROSE K., IIDA S.: A proof of negative answer to Hilbert's 10th problem, *Proc. Japan Acad.* Vol. 49, 1973 pp. 10-12
- HIRSCHFELD J.: Existentially complete and generic structures in arithmetic. In: *Dissertation*, Yale, 1972
- HIRSCHFELD J.: Models of arithmetic and the semiring of recursive functions. In: *Victoria Symposium on Non-Standard analysis* (Lect. Notes Math. 369), Springer-Verlag, 1974 pp. 99-105
- HIRSCHFELD J.: Models of arithmetic and recursive functions, *Israel J. Math.* Vol. 20, 1975 pp. 111-126
- HIRSCHFELD J.: Finite forcing and generic filters in arithmetic. In: *Model Theory and Algebra, A Memorial Tribute to Abraham Robinson*, Springer-Verlag, 1975, pp. 172-199
- HIRSCHFELD J., WHEELER W. H.: Forcing, arithmetic, division rings. In: (Lect. Notes Math. 454), Springer-Verlag, 1975, 266 p.
- HODGES W.: Interpreting number theory in nilpotent groups, *Arch. Math. Log.* Vol. 20, 1980 pp. 103-111
- HODGES W.: Truth in a structure. In: *Proc. of the Aristotelian Society* (New series) vol. 86, 1985/86, 135-151

- HOPCROFT J. E., PAUL W., VALIANT L.: On time versus space and related problems, 16-th Symp. on Foundations of Comp. Sci., 1975 pp. 57-64
- HÖRING W.: Absolut unentscheidbare Sätze der Mathematik. In: *Logik und Logikkalkül*, Alber: Freiburg, 1962 pp. 189-194
- HOWARD P. E.: A proof of the theorem of Tennenbaum, *Zeitschr. Math. Log. Grundl. Math.* Vol. 18, 1972 pp. 111-112
- HUBER-DYSON V., JONES J. P., SHEPERDSON J. C.: Some diophantine forms of Gödel's theorem, *Arch. Math. Log.* Vol. 22, 1982 pp. 51-60
- IGNJATOVIC A.: An ordering of the set of sentences of Peano arithmetic, *Publ. Inst. Math, Nouv. Ser.* Vol. 38 (52), 1985 pp. 13-15
- IMMERMAN N.: Nondeterministic space is closed under complement. In: *Proc. 3-rd Annual Conf. Structure in Complexity Theory*, IEEE, 1988 pp. 112-115
- ISAACSON D.: Arithmetical truth and hidden higher-order concepts. In: *Logic Colloquium '85*, Amsterdam: North-Holland, 1987 pp. 147-169
- JENSEN D.: On local proof restrictions for strong theories, *Diss. Math.* Vol. 108, 1973 pp. 1-47
- JENSEN D. EHRENFEUCHT A.: Some problem in elementary arithmetics, *Fund. Math.* Vol. 92, 1976 pp. 223-245
- JENSEN R. B.: Unabhängigkeitsbeweise in Teilsystemen der elementaren Zahlentheorie, *Math. Ann.* Vol. 173, 1967 pp. 102-126
- JEROSLOW R. G.: Non-effectiveness in S. Orey's compactness theorem, *Zeitschr. Math. Log. Grundl. Math.* Vol. 17, 1971 pp. 285-289
- JEROSLOW R. G.: Consistency statements in formal theories, *Fund. Math.* Vol. 72, 1971 pp. 17-40
- JEROSLOW R. G.: Redundancies in the Hilbert-Bernays derivability conditions for Gödel's second incompleteness theorem, *Journ. Symb. Log.* Vol. 38, 1973 pp. 359-367
- JEROSLOW R. G.: Experimental logics and Δ_2^0 -theories, *Journ. Phil. Log.* Vol. 4, 1975 pp. 253-267
- JERVEL H. R.: Gentzen games. *Z Math Logik* Vol. 31, pp. 431-439
- JERVEL H. R.: Large finite sets, *Z Math Logik* Vol. 31, pp. 545-549
- JOCKUSCH C. G.: Ramsey's theorem and recursion theory, *Journ. Symb. Log.* Vol. 37, 1972 pp. 268-280
- JOCKUSCH C. G.: Π_1^0 classes and boolean combinations of recursively enumerable sets, *Journ. Symb. Log.* Vol. 39, 1974 pp. 95-96
- JOCKUSCH C. G., SOARE R. I.: Π_1^0 -classes and degrees of theories, *Trans. Amer. Math. Soc.* Vol. 173, 1972 pp. 33-56
- JOCKUSCH C. G., SOARE R. I.: Degrees of members of Π_1^0 classes, *Pac. Journ. Math.* Vol. 40, 1972 pp. 605-616
- JONES J. P.: Formula for the n -th prime number, *Can. Math. Bull.* Vol. 18, 1975 pp. 433-434
- JONES J. P.: Diophantine representation of the Fibonacci numbers, *Fib. Quart.* Vol. 13, 1975 pp. 82-88
- JONES J. P.: Diophantine representation of the Lucas numbers, *Fib. Quart.* Vol. 14, 1976 pp. 134
- JONES J. P.: Three universal representation of recursively enumerable sets, *Journ. Symb. Log.* Vol. 43, 1978 pp. 335-351
- JONES J. P.: Diophantine representation of Mersenne and Fermat primes, *Acta Arithmetica* Vol. 35, 1979 pp. 209-221
- JONES J. P.: Udecidable Diophantine equations, *Bull. Amer. Math. Soc. (NS)* Vol. 3, 1980 pp. 859-862
- JONES J. P.: Universal diophantine equation, *Journ. Symb. Log.* Vol. 47, 1982 pp. 549-571
- JONES J. P., LEVITZ H., WILKIE A. J.: Classification of quantifier prefixes over exponential diophantine equations, *Zeitschr. Math. Log. Grundl. Math.* Vol. 32, 1986 pp. 399-406

- JONES J. P., MATIYASEVIČ YU. V.: Register machine proof of the theorem on exponential diophantine representation of enumerable sets, *Journ. Symb. Log.* Vol. 49, 1984 pp. 48–59
- JONES J. P., SATO D., WADA H., WIENS D.: Diophantine representation of the set of prime numbers, *Amer. Math. Monthly* Vol. 83, 1976 pp. 449–464
- JONES J. P., SHEPHERDSON J. C.: Variants of Robinson's essentially undecidable theory R, *Arch. Math. Log.* Vol. 23, 1983 pp. 61–64
- DE JONGH D.: A simplification of a completeness proof of Guaspari and Solovay, *Studia logica* Vol. 46, 1987 pp. 187–192
- DE JONGH D.: Provability logics for relative interpretability (preprint for the Heyting conference Varna), 1988, 19 p.
- DE JONGH D., MONTAGNA F.: Provable fixed points, *Zeitschr. Math. Log. Grundle. Math.* Vol. 34, 1988 pp. 229–250
- DE JONGH D., MONTAGNA F.: Much shorter proofs, Report 87-13, Dept. Math. Univ. Amsterdam, 1987, 21 p.
- DE JONGH D., MONTAGNA F.: Generic generalized Rosser fixed points, *Studia logica* Vol. 46, 1987 pp. 193–203
- DE JONGH D., MONTAGNA F.: Rosser orderings and free variables (preprint for Heyting conference Varna), 1988, 10 p.
- DE JONGH D., MONTAGNA F.: Provable fixed points, *Zeitschr. Math. Log. Grundle. Math.* Vol. 34, 1988 pp. 229–250
- JOSEPH D., YOUNG P.: Independence results in computer science, *Journ. Comp. Syst.* Vol. 23, 1981 pp. 205–222
- JOSEPH D., YOUNG P.: A survey of some recent results on computational complexity in weak theories of arithmetic. In: *Mathematical foundations of computer science (Lect. Notes. Comp. Sci 118)*, Springer-Verlag, 1981 pp. 46–60
- JUMELET M.: On Solovay's completeness theorem (thesis), Amsterdam, 1988, 28 p.
- KALSBECK M. B.: An Orey sentence for predicative arithmetic (master's thesis), Amsterdam 1988
- KAMO S.: Nonstandard natural number systems and nonstandard models, *Journ. Symb. Log.* Vol. 46, 1981 pp. 365–376
- KANAMORI A., MCALOON K.: On Gödel incompleteness and finite combinatorics, *Ann. Pure Appl. Logic* Vol. 33, 1987 pp. 23–41
- KANOVIC M. I.: An estimate of complexity of arithmetic incompleteness, *Doklady Akad. Nauk SSSR* Vol. 238, 1978 pp. 1283–1286
- KANOVIC M. I.: Nezavisimost invariantnykh predlozenij, *Doklady Akad. Nauk SSSR*, 1984 No. 276 pp. 27–31
- KARP R. M.: Reducibility among combinatorial problems, in *Complexity of Computer Computations*, Miller and Thatcher eds., Plenum Press, New York 1972 pp. 85–103
- KARP R. M., LIPTON R. J.: Some connections between nonuniform and uniform complexity classes. In: *Proc. 12-th Annual ACM Symp. on Theory of Computing*, 1980, pp. 302–309
- KAUFMANN M.: A rather classless model, *Proc. Amer. Math. Soc.* Vol. 62, 1977 pp. 330–333
- KAUFMANN M.: On existence Σ_n -end extensions. In: *Logic Year 1979–80. The University of Connecticut (Lect. Notes Math. 859)*, Springer-Verlag, 1981 pp. 92–103
- KAUFMANN M.: On expandability of models of arithmetic and set theory to models of weak second-order theories, *Fund. Math.* Vol. 122, 1984 pp. 57–60
- KAUFMANN M., SCHMERL J. H.: Saturation and simple extensions of models of Peano arithmetic, *Annals Pure Appl. Logic* Vol. 27, 1984 pp. 109–136
- KAUFMANN M., SCHMERL J. H.: Remarks on weak notions of saturation in models of Peano arithmetic, *Journ. Symb. Log.* Vol. 52, 1987 pp. 129–148
- KAYE R.: Characterizing some low complexity classes using weak theories of arithmetic, Transfer Report, Manchester University, 1985

- KAYE R.: Parameter free induction. In: Proc. of the 5th Easter conf. on model theory (Humboldt univ. Berlin), 1987
- KAYE R.: Diophantine and parameter-free induction (thesis), Manchester University, 1987
- KAYE R.: Axiomatizations and quantifier complexity. In: Proc. of the 6th Easter conf. on model theory (Humboldt Univ. Berlin), 1988, pp. 65–84
- KAYE R.: On functions provably E_1 in open induction, 1985 (manuscript)
- KAYE R.: Parameter-free universal induction, *Zeitschr. f. Math. Logik* 35, 1989, pp. 443–456
- KAYE R.: Diophantine induction, *Annals of Pure and Appl. Logic* 46, 1990, pp. 1–40
- KAYE R.: Models of Peano arithmetic, *Oxford Logic Guides*, Oxford Univ. Press 1991, 292 pp.
- KAYE R.: Model-theoretic properties characterizing Peano arithmetic, *Journ. Symb. Logic* 56, 1991, pp. 949–963
- KAYE R.: On cofinal extensions of models of fragments of arithmetic, *Notre Dame J. of Formal Logic* 32, 1991, pp. 399–408
- KAYE R.: Open induction, Tennenbaum phenomena, and complexity theory. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajčček, Oxford Univ. Press 1992, pp. 222–237
- KAYE R.: Using Herbrand-type theorems to separate, strong fragments of arithmetic. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajčček, Oxford Univ. Press 1992, pp. 238–246
- KAYE R., KOSSAK R., KOTLARSKI H.: Automorphisms of recursively saturated models of arithmetic, *Annals of Pure and Appl. Logic* 55, 1991, pp. 67–99
- KAYE R., PARIS J. B., DIMITRACOPOULOS C.: On parameter free induction schemas, *Journ. of Symb. Logic* 53, 1988, pp. 1082–1097
- KELLER J. P., RICHARD D.: Remarques sur les structures additives des modèles de l'arithmétique, *Comptes Rendus, Sér. A* Vol. 287, 1978 pp. 101–104
- KEMENY J. G.: Undecidable problems of elementary number theory, *Math. Ann.* Vol. 135, 1958 pp. 160–169
- KENT C. F.: The relation of A to prov 'A' in the Lindenbaum sentence algebra, *Journ. Symb. Log.* Vol. 38, 1973 pp. 295–298
- KENT C. F.: "Disorder" in lattices of binumerations, *Comm. Math. Univ. Carol.* Vol. 15, 1974 pp. 221–244
- KENT C. F.: Independence versus logical independence in the countable case. In: *Logic Colloquium '73*, North Holland: Amsterdam, 1975, pp. 399–408
- KETONEN J., SOLOVAY R. M.: Rapidly growing Ramsey functions, *Ann. Math.* Vol. 113, 1981 pp. 267–314
- KINO A.: On provably recursive functions and ordinal recursive functions, *Journ. Math. Soc. Japan* Vol. 20, 1968 pp. 456–476
- KIRBY L. A. S.: Initial segments of models of arithmetic. In: *Dissertation*, Manchester, 1977
- KIRBY L. A. S.: La méthode des indicatrices et le théorème d'incomplétude. In: *Models of arithmetic* (Sem. Univ. Paris VII, Paris, 1977) (French), Soc. Math. France, Paris, 1980 pp. 5–18
- KIRBY L. A. S.: Flipping properties in arithmetic, *Journ. Symb. Log.* Vol. 47, 1982 pp. 416–422
- KIRBY L. A. S.: Ultrafilters and types on models of arithmetic, *Annals Pure Appl. Logic* Vol. 27, 1984 pp. 215–252
- KIRBY L. A. S., MCALOON K., MURAWSKI R.: Indicators, recursive saturation and expandability, *Fund. Math.* Vol. 114, 1981 pp. 127–139
- KIRBY L. A. S., PARIS J. B.: Initial segments of models of Peano's axioms. In: *Set Theory and Hierarchy Theory V. Proceedings of the 3rd Conference on Set Theory and Hierarchy Theory*, Springer-Verlag, 1977 pp. 211–226

- KIRBY L. A. S., PARIS J. B.: Σ_n collection schemas in arithmetic. In: *Logic Colloquium 77* (Stud Logicae Fundam Math 96), North Holland: Amsterdam, 1978 pp. 199–209
- KIRBY L. A. S., PARIS J. B.: Accessible independence results for Peano arithmetic, *Bull. London Math. Soc.* Vol. 14, 1982 pp. 285–293
- KISS P.: Diophantine representation of generalized Fibonacci numbers, *Elemente der Math.* Vol. 34, 1979 pp. 129–132
- KLEENE S.: Recursive functions and predicates, *Trans. Amer. Math. Soc.* Vol. 53, 1943 pp. 41–73
- KLEENE S. C.: A symmetric form of Gödel's numbers, *Indag. Math.* Vol. 12, 1950 pp. 244–246
- KLEENE S. C.: Finite axiomatizability of theories in the predicate calculus using additional predicate symbols. *Memoirs of the Amer. Math. Soc.*, 1951 pp. 27–66 (Russian translation in *Matematicheskaja teorija logičeskogo vyvoda*, Nauka 1967, 237–284)
- KLEENE S. C.: On the forms of predicates in the theory of constructive ordinals I, *Amer. Journ. of Math.* vol. 66, 1944, pp. 41–58
- KLEENE S. C.: *Introduction to metamathematics*, North Holland, Amsterdam, 1952, 550 p.
- KLEENE S. C.: On the forms of predicates on the theory of constructive ordinals II, *Amer. Journ. of Math.* vol 77, 1955, pp. 405–428
- KLEENE S. C.: *Mathematical logic*, Wiley & Sons: New York, 1967, 398 p.
- KLINGEN N.: Zur Idealstruktur in Nichtstandardmodellen von Dedekindringen, *Journ. Reine Angew. Math* Vol. 274/275, 1975 pp. 38–60
- KNIGHT J. F.: Complete types and the natural numbers, *Journ. Symb. Log.* Vol. 38, 1973 pp. 413–415
- KNIGHT J. F.: Types omitted in uncountable models of arithmetic, *Journ. Symb. Log.* Vol. 40, 1975 pp. 317–320
- KNIGHT J. F.: Omitting types in set theory and arithmetic, *Journ. Symb. Log.* Vol. 41, 1976 pp. 25–32
- KNIGHT J. F.: Hanf numbers for omitting types over particular theories, *Journ. Symb. Log.* Vol. 41, 1976 pp. 583–588
- KNIGHT J. F.: Models of arithmetic and closed ideals, *Journ. Symb. Log.* Vol. 47, 1982 pp. 833
- KNIGHT J. F.: Additive structure in uncountable models for a fixed completion of P, *Journ. Symb. Log.* Vol. 48, 1983 pp. 623–628
- KNIGHT J. F.: Effective construction of models. In: *Logic Colloquium '84* (Studies in Logic and the Foundations of Mathematics 120), North-Holland, 1986 pp. 105–119
- KNIGHT J. F., LACHLAN A. H., SOARE R. I.: Two theorems on degrees of models of true arithmetic, *Journ. Symb. Log.* Vol. 49, 1984 pp. 425–436
- KNIGHT J. F., NADEL M.: Expansions of models and Turing degrees, *Journ. Symb. Log.* Vol. 47, 1982 pp. 587
- KOCHEN S., KRIPKE S. A.: Nonstandard models of Peano arithmetic, *Enseign. Math.*, Ser 2 Vol. 28, 1982 pp. 211–231
- KÖNIG D.: Über eine Schlußweise aus dem Endlichen ins Unendliche, *Acta litterarum ac scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio scientiarum mathematicarum* vol. 3, 1927, pp. 121–130
- KOPPEL M.: Some decidable Diophantine problems: positive solution to a problem of Davis, Matiyasevič and Robinson, *Proc. Amer. Math. Soc.* Vol. 77, 1979 pp. 319–323
- KOSSAK R.: An application of definable types of Peano's arithmetic, *Bull. Acad. Polon. Sci.* Vol. 28, 1980 pp. 213–217
- KOSSAK R.: A certain class of models of Peano arithmetic, *Journ. Symb. Log.* Vol. 48, 1983 pp. 311–320
- KOSSAK R.: L_{∞, ω_1} -elementary equivalence of ω_1 -like models of PA, *Fund. Math.* Vol. 123, 1984 pp. 123–131

- KOSSAK R.: A note on satisfaction classes, *Notre Dame Journ. Formal Log.* Vol. 26, 1985 pp. 1-8
- KOSSAK R.: Recursively saturated ω_1 -like models of arithmetic, *Notre Dame Journ. Formal Log.* Vol. 26, 1985 pp. 413-422
- KOSSAK R.: Remarks on free sets, *Bull. Pol. Acad. Sci.* Vol. 34, 1986 pp. 117-122
- KOSSAK R.: Models with the ω property, *Journ. Symb. Log.* Vol. 54, 1989 pp. 177-189
- KOSSAK R., KOTLARSKI H.: Results on automorphisms of recursively saturated models of PA, *Fund. Math.* Vol. 129, 1988 pp. 1-7
- KOSSAK R., PARIS J. B.: Subsets of models of arithmetic. In: *Open Days in Model Th. & Set Th.*; 1981 *Jadwisin*, 1981 pp. 159-174
- KOSSOVSKY N. K.: Diophantine representation of the sequence of solutions to the Pell equation, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 20, 1971 pp. 28-35
- KOTLARSKI H.: On Skolem ultrapowers and their non-standard variant, *Zeitschr. Math. Log. Grundle. Math.* Vol. 26, 1980 pp. 227-236
- KOTLARSKI H.: On elementary cuts in models of arithmetic, *Bull. Acad. Polon. Sci.* Vol. 29, 1981 pp. 419-423
- KOTLARSKI H.: On confinal extensions of models of arithmetic, *Journ. Symb. Log.* Vol. 48, 1983 pp. 253-262
- KOTLARSKI H.: On elementary cuts in models of arithmetic, *Fund. Math.* Vol. 115, 1983 pp. 27-31
- KOTLARSKI H.: On elementary cuts in recursively saturated models of Peano arithmetic, *Fund. Math.* Vol. 120, 1984 pp. 205-222
- KOTLARSKI H.: Some remarks on initial segments in models of Peano arithmetic, *Journ. Symb. Log.* Vol. 49, 1984 pp. 955-960
- KOTLARSKI H.: Bounded induction and satisfaction classes. In: *Proceedings of the 3rd Easter Conference on Model Theory*, Humboldt-Univ. Berlin: Berlin, 1985 pp. 143-167
- KOTLARSKI H.: The recursively saturated part of models of Peano arithmetic, *Zeitschr. Math. Log. Grundle. Math.* Vol. 32, 1986 pp. 365-370
- KOTLARSKI H.: Bounded induction and satisfaction classes, *Zeitschr. Math. Log. Grundle. Math.* Vol. 32, 1986 pp. 531-544
- KOTLARSKI H.: On the end extension problem for Δ_0 -PA(S), *Zeitschr. f. Math. Logik* 35, 1989, pp. 391-397
- KOTLARSKI H., KRAJEWSKI S., LACHLAN A. H.: Construction of satisfaction classes for nonstandard models, *Canad Math Bull* Vol. 24, 1981 pp. 283-293
- KOTLARSKI H., RATAJCZYK Z.: A weak extension of bounded induction in the language with a satisfaction class, *Journ. Symb. Log.* Vol. 52, 1987 pp. 1070
- KOTLARSKI H., RATAJCZYK Z.: Inductive full satisfaction classes, *Annals Pure Appl. Logic* Vol. 47, 1990, pp. 199-223
- KRAJEWSKI S.: Predicative expansions of axiomatic theories, *Zeitschr. Math. Log. Grundle. Math.* Vol. 20, 1974 pp. 435-452
- KRAJEWSKI S.: Mutually inconsistent satisfaction classes, *Bull. Acad. Polon. Sci.* Vol. 22, 1974 pp. 983-987
- KRAJEWSKI S.: A remark on automorphisms and nonstandard properties, *Bull. Acad. Polon. Sci.* Vol. 22, 1974 pp. 989-991
- KRAJEWSKI S.: Non-standard satisfaction classes. In: *Set Theory and Hierarchy Theory. A Memorial Tribute to Andrzej Mostowski. Proceedings of the 2nd Conference on Set Theory and Hierarchy Theory* (Lect. Notes. Math 537), Springer-Verlag, 1976 pp. 121-144
- KRAJEWSKI S.: A note on expansion of models of set theories. In: *Set Theory and Hierarchy Theory. Proceedings of the 1st Colloquium in Set Theory and Hierarchy Theory*, Politechnical University Wroclaw, 1977 pp. 63-67
- KRAJÍČEK J.: Generalizations of proofs. In: *Proc. 8-th Easter Conference on Model Theory*, Wendisch-Rietz, 1987 pp. 82-99

- KRAJÍČEK J.: A note on proofs of falsehood, *Arch. Math. Log.* Vol. 26, 1987 pp. 169–176
- KRAJÍČEK J.: Speed-up for propositional Frege systems via generalizations of proofs, *Comm. Math. Univ. Carol.* Vol. 30, 1989, pp. 137–140
- KRAJÍČEK J.: On the number of steps in proofs, *Ann. Pure Appl. Logic*, Vol. 41, 1989, pp. 153–178
- KRAJÍČEK J.: A theorem on uniform provability of schemes. In: *Proc. 6-th Easter Conf. on Model Theory*, Wendisch-Rietz, 1988 pp. 85–92
- KRAJÍČEK J.: Exponentiation and second order bounded arithmetic, *Ann. Pure Appl. Logic*, Vol. 48, 1990, pp. 261–276
- KRAJÍČEK J.: No counter-example interpretation and interactive computation. In: *Logic from Computer Science*, ed. Y. N. Moschovakis, Springer-Verlag 1992, pp. 287–293
- KRAJÍČEK J.: Fragments of Bounded Arithmetic and bounded query classes, *Transactions AMS*, to appear
- KRAJÍČEK J., PUDLÁK P.: The number of proof lines and the size of proofs in first order logic, *Arch. Math. Log.* Vol. 27, 1988 pp. 69–84
- KRAJÍČEK J., PUDLÁK P.: On the structure of initial segments of models of arithmetic, *Arch. Math. Log.* Vol. 28, 1989 pp. 91–98
- KRAJÍČEK J., PUDLÁK P.: Propositional provability and models of weak arithmetic. In: *Proc. Computer Science Logic '89*, (Lect. Notes Comp. Sci. 440), Springer-Verlag, 1990, pp. 193–210
- KRAJÍČEK J., PUDLÁK P.: Propositional proof systems, the consistency of first order theories and the complexity of computations, *Journ. Symb. Log.* Vol. 54, 1989 pp. 1063–1079
- KRAJÍČEK J., PUDLÁK P.: Quantified propositional calculi and fragments of bounded arithmetic, *Zeitschr. Math. Log. Grundle. Math.* Vol. 36, 1989 pp. 29–46
- KRAJÍČEK J., PUDLÁK P., TAKEUTI G.: Bounded Arithmetic and the Polynomial Hierarchy, *Annals of Pure and Appl. Logic* 52, pp. 143–154
- KRAJÍČEK J., TAKEUTI G.: On induction-free provability, *Annals of Mathematics and Artificial Intelligence*, (to appear)
- KRAJÍČEK J., TAKEUTI G.: On bounded Σ_1^1 polynomial induction. In: *Feasible Mathematics*, Buss, Scott eds., Birkhäuser 1990, pp. 259–280
- KREISEL G.: A survey of proof theory II. In: *Proceedings of the Second Scandinavian Logic Symposium*, North Holland: Amsterdam pp. 109–170
- KREISEL G.: Informal rigour and completeness proofs. In: *Proceedings of the International Colloquium in the Philosophy of Science* Vol. 1, North-Holland: Amsterdam, 1967, pp. 138–186
- KREISEL G.: Note on arithmetic models for consistent formulae of the predicate calculus I, *Fund. Math.* Vol. 37, 1950 pp. 265–285
- KREISEL G.: On the interpretation of non-finitist proofs I, *Journ. Symb. Logic* Vol. 16, 1951 pp. 241–267
- KREISEL G.: On the concepts of completeness and interpretation of formal systems, *Fund. Math.* Vol. 39, 1952 pp. 103–127
- KREISEL G.: Some concepts concerning formal systems of number theory, *Zeitschr. Math. Log. Grundle. Math.* Vol. 57, 1952 pp. 1–12
- KREISEL G.: On the interpretation of non-finitist proofs II. Interpretation of number theory. Applications, *Journ. Symb. Log.* Vol. 17, 1952 pp. 43–58
- KREISEL G.: The diagonal method in formalized arithmetic, *Brit. Journ. Phil. Sci.* Vol. 3, 1953 pp. 364–373
- KREISEL G.: On a problem of Henkin's, *Indag. Math.* Vol. 15, 1953 pp. 405–406
- KREISEL G.: Note on arithmetic models for consistent formulae of the predicate calculus II. In: *Actes du 11eme Congres International de Philosophie*, North Holland: Amsterdam Vol. 14, 1953 pp. 39–49
- KREISEL G.: A variant to Hilbert's theory of the foundations of arithmetic, *Br. J. Phil. Sci.* Vol. 4, 1953 pp. 107–129

- KREISEL G.: Remark on complete interpretations by models, *Arch. Math. Log.* Vol. 2, 1954 pp. 4-9
- KREISEL G.: Models, translations and interpretations. In: *Mathematical Interpretation of Formal Systems*; 1954 Amsterdam (*Stud Logic Found Math* 10), North Holland, Amsterdam, 1955 pp. 26-50
- KREISEL G.: Gödel's interpretation of Heyting's arithmetic. In: *Summaries of Talks Presented at the Summer Institute for Symbolic Logic, Institute for Defense Analyses, Comm. Res. Division*, 1957 pp. 125-133
- KREISEL G.: On weak completeness of intuitionistic predicate logic, *Journ. Symb. Log.* Vol. 27, 1962 pp. 139-158
- KREISEL G.: Foundations of intuitionistic logic. In: *Logic, Methodology and Philosophy of Science*, Stanford, 1962, pp. 198-210
- KREISEL G.: Mathematical logic. In: *Lect. Modern Math.*, III, Wiley, New York, 1965, pp. 95-195
- KREISEL G.: A survey on proof theory, *Journ. Symb. Log.* Vol. 33, 1968 pp. 321-388
- KREISEL G.: Axiomatizations of nonstandard analysis that are conservative extensions of formal systems for classical standard analysis. In: *Applications of Model Theory to Algebra, Analysis and Probability*, Holt, Rinehart and Winston, 1969 pp. 93-106
- KREISEL G.: Which number theoretic problems can be solved in recursive progressions on Π_1^1 -paths through O ? *Journ. Symb. Log.* Vol. 37, 1972 pp. 311-334
- KREISEL G.: Observations on a recent generalization of completeness theorems due to Schütte. In: *Proof Theory Symposium, Kiel 1974 (Lect. Notes Math. 500)*, Springer-Verlag, 1975 pp. 164-181
- KREISEL G.: What have we learnt from Hilbert's second problem? In: *Mathematical Developments Arising From Hilbert's Problems*, Amer. Math. Soc.: Providence, 1976, pp. 93-130
- KREISEL G.: Wie die Beweistheorie zu ihren Ordinalzahlen kam und kommt, *Jber. d. Dt. Math.-Verein.* Vol. 78, 1977 pp. 177-223
- KREISEL G.: Finiteness theorems in arithmetic: an application of Herbrand's theorem for Σ_2 formulas. In: *Proceedings of the Herbrand Symposium. Logic Colloquium '81 (Stud Logic Found Math 107)*, North Holland: Amsterdam, 1982 pp. 39-55
- KREISEL G., LEVY A.: Reflections principles and their use for establishing the complexity of axiomatic systems, *Zeitschr. Math. Log. Grundle. Math.* Vol. 14, 1968 pp. 97-142
- KREISEL G., MINC G. E. SIMPSON S. G.: The use of abstract languages in elementary mathematics: some pedagogic examples. In: *Logic Colloquium*, Springer-Verlag, 1975 pp. 38-131
- KREISEL G., SHOENFIELD J. R., WANG H.: Number theoretic concepts and recursive well-orderings, *Arch. Math. Log.* Vol. 5, 1959 pp. 42-64
- KREISEL G., TAKEUTI G.: Formally self-referential propositions in cut-free classical analysis and related systems. *Diss. Math.* Vol. 118, 1974 pp. 1-50
- KREISEL G., WANG H.: Some applications of formalized consistency proofs, *Fund. Math.* Vol. 42 (dod. 45), 1955 pp. 101n. (334)
- KRIAUCIUKAS V.: A Diophantine representation of perfect numbers, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 88, 1979 pp. 78-89
- KRIPKE S. A.: "Flexible" predicates of formal number theory, *Proc. Amer. Math. Soc.* Vol. 13, 1962 pp. 647-650
- KRIPKE S. A., POUR-EL M. B.: Deduction preserving recursive isomorphisms of theories, *Fund. Math.* Vol. 61, 1967 pp. 141-147
- KUČERA A.: Measure, Π_1^0 -classes and complete extensions of PA. In: *Recursion Theory Week (Lect. Notes Math. 1141)*, Springer-Verlag, 1985 pp. 245-259
- KUČERA A.: Alternative, priority-free constructions of some r. e. degrees and other results on degrees, *Journ. Symb. Log.* Vol. 52, 1987 pp. 1071

- KURATA R.: The reflection principle, transfinite induction, and the Paris-Harrington principle. In: *Boolean-Algebra-Valued Analysis and Nonstandard Analysis*, Kyoto Univ Res Inst Math Sci: Kyoto, 1981 pp. 1-14
- KURATA R.: Paris-Harrington theory and reflection principles, *Saitama Math. J.* Vol. 2, 1984 pp. 33-45
- KURATA R.: A simple proof for a statement which is equivalent to Harrington's principle. (abstract), *Journ. Symb. Log.* Vol. 50, 1985 pp. 268
- KURATA R.: Paris-Harrington theory and principles. (abstract), *Journ. Symb. Log.* Vol. 51, 1986 pp. 489
- KURATA R.: Paris-Harrington principles, reflection principles and transfinite induction up to ϵ_0 , *Annals Pure Appl. Logic* Vol. 31, 1986 pp. 237-256
- LACHLAN A. H.: Full satisfaction classes and recursive saturation, *Canad. Math. Bull.* Vol. 24, 1981 pp. 295-297
- LADNER R. E.: On the structure of polynomial time reducibility, *Journ. of ACM* Vol. 22, 1975 pp. 155-171
- LASCAR D.: Une indicatrice de type "Ramsey" pour l'arithmétique de Peano et la formule de Paris-Harrington. In: *Models of arithmetic* (Sem, Univ. Paris VII, Paris, 1977), Soc. Math. France, Paris, 1980 pp. 19-30
- VAN LEEUWEN J. (ED.): *Handbook of Theoretical Computer Science*, Vol. A - Algorithms and Complexity, North-Holland, 1990
- LEIVANT D.: On the proof theory of the modal logic for arithmetic provability, *Journ. Symb. Log.* Vol. 46, 1981 pp. 531-538
- LESAN H.: *Models of arithmetic*. Dissertation, Manchester, 1978
- LEVITZ H.: An ordered set of arithmetic functions representing the least epsilon-number, *Zeitschr. Math. Log. Grundle. Math.* Vol. 21, 1975 pp. 115-120
- LEVITZ H.: An initial segment of the set of polynomial functions with exponentiation, *Alg. Universalis* Vol. 7, 1977 pp. 133-136
- LEVITZ H.: An ordinal bound for the set of polynomials with exponentiation, *Alg. Universalis* Vol. 8, 1978 pp. 233-243
- LINDSTRÖM P.: Some results on relative interpretability, *Phil. communications, red series*, No. 7, Univ. of Goteborg, 1978
- LINDSTRÖM P.: Some results on interpretability. In: *Proc. 5th Scand. Symp.* (Jensen et al, ed.), Aalborg, 1979 pp. 329-361
- LINDSTRÖM P.: Notes on partially conservative sentences and interpretability, *Phil. communications, red series*, No. 13, Univ. of Goteborg, 1980
- LINDSTRÖM P.: More on partially conservative sentences and interpretability, *Proc. Amer. Math. Soc.* Vol. 91, 1984 pp. 436-443
- LINDSTRÖM P.: On certain lattices of degrees of interpretability, *Notre Dame Journ. Formal Log.* Vol. 25, 1984 pp. 127-140
- LINDSTRÖM P.: On faithful interpretability. In: *Computation and proof theory*, *Proc. Logic Colloq, Aachen 1983*, Part II, (Lect. Notes Math. 1104), Springer-Verlag, 1984 pp. 279-288
- LINDSTRÖM P.: On partially conservative sentences and interpretability, *Proc. Amer. Math. Soc.* Vol. 91, 1984 pp. 436-444
- LINDSTRÖM P.: Notes on formulas with prescribed properties in arithmetical theories, *Philosophical Communications, red series* No. 25, Univ. of Goteborg, 1984
- LINDSTRÖM P.: Provability and interpretability in theories containing arithmetic. In: *Atti degli incontri di logica matematica*, Vol. 2, Siena / Italia 1983/84, 1985 pp. 431-451
- LINDSTRÖM P.: Partially generic formulas in arithmetic, *Notre Dame Journ. Formal Log.* Vol. 29, 1988 pp. 185-192
- LIPSHITZ L.: Undecidable existential problems for addition and divisibility in algebraic number rings II, *Proc. Amer. Math. Soc.* Vol. 64, 1977 pp. 122-128
- LIPSHITZ L.: Undecidable existential problems for addition and divisibility in algebraic number rings, *Trans. Amer. Math. Soc.* Vol. 241, 1978 pp. 121-128

- LIPSHITZ L.: The Diophantine problem for addition and divisibility, *Trans. Amer. Math. Soc.* Vol. 235, 1978 pp. 271–284
- LIPSHITZ L.: Diophantine correct models of arithmetic. *Proc. Amer. Math. Soc.* Vol. 73, 1979 pp. 107–108
- LIPSHITZ L., NADEL M. E.: The additive structure of models of arithmetic, *Proc. Amer. Math. Soc.* Vol. 68, 1978 pp. 331–336
- LIPTON R. J.: On consistency of $P=NP$ and fragments of arithmetic. In: *Fundamentals of Computation Theory – FCT'79*, Akademie Verlag: Berlin, 1979 pp. 269–278
- LOEB M. H.: Solution of a problem by Leon Henkin. In: *Proceedings of the International Congress of Mathematicians 1954*, North Holland, Amsterdam Vol. 2, 1954 pp. 405–406
- LOEB M. H.: Solution of a problem of Leon Henkin, *Journ. Symb. Log.* Vol. 20, 1955 pp. 115–118
- LOEB M. H., WAINER S. S.: Hierarchies of number-theoretic functions I,II, *Arch. Math. Log.* vol. 13, 1970, pp. 39–51, 97–113 (a correction on vol. 14, 1971, 198–199)
- LOEBL M.: Hercules and Hydra, a game on rooted finite trees, *Comm. Math. Univ. Carol.* Vol. 26, 1985 pp. 259–267
- LOPEZ-ESCOBAR E. G. K.: König's lemma, the ω -rule and primitive recursive arithmetic, *Arch. Math. Log.* Vol. 25, 1985 pp. 67–74
- ŁOŚ J.: On the extending of models I, *Fund. Math.* vol. 42, 1955, 38–54
- MACDOWELL R., SPECKER E.: Modelle der Arithmetik. In: *Infinitistic Methods. Proceedings of the Symposium on Foundations of Mathematics*, PWN: Warsaw, 1961 pp. 257–263
- MACINTYRE A.: Ramsey quantifiers in arithmetic. In: *Model theory of algebra and arithmetic (Proc. Conf, Karpacz, 1979)* (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 186–210
- MACINTYRE A.: Nonstandard number theory. In: *Proceedings of the International Congress of Mathematicians, Academia Scientiarum Fennica: Helsinki* Vol. 1, 1980 pp. 253–262
- MACINTYRE A.: The laws of exponentiation. In: *Model theory and arithmetic (Paris, 1979–1980)* (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 185–197
- MACINTYRE A.: A theorem of Rabin in a general setting. *Bull. Soc. Math. Belg., Ser B* Vol. 33, 1981 pp. 53–63
- MACINTYRE A.: Residue fields of models of P . In: *Proceedings of the 6th International Congress for Logic, Methodology and Philosophy of Science (Stud Logic Found Math 104)*, North Holland: Amsterdam, 1982 pp. 193–206
- MACINTYRE A.: The strength of weak systems. In: *Schriftenreihe der Wittgenstein-Gesellschaft, 13 Logic, Philosophy of Science and Epistemology*, Wien, 1987 pp. 43–59
- MACINTYRE A., MARKER D.: Degrees of recursively saturated models, *Trans. Amer. Math. Soc.* Vol. 282, 1984 pp. 539–554
- MACINTYRE A., MARKER D.: Primes and their residue rings in models of open induction, *Annals of Pure and Appl. Log.* Vol. 43, 1989, pp. 57–77
- MACINTYRE A., SIMMONS H.: Gödel's diagonalization technique and related properties of theories. In: *Colloquium Mathematicum, Academie Polonaise des Sciences, Inst Math: Warsaw* Vol. 28, 1973 pp. 165–180, 329
- MACINTYRE A., SIMMONS H.: Algebraic properties of number theories, *Israel Journ. Math.* Vol. 22, 1975 pp. 7–27
- MAGARI R.: Meaning and truth in the Peano arithmetic, *Atti della Accademia Nazionale dei Lincei. Rendiconti. Classe di Scienze Fisiche, Matematiche e Naturali. Serie VIII* Vol. 54, 1973 pp. 902–903
- MAGARI R.: Significate e verita nell' aritmetice peaniane, *Annali di Mat. pure ed applicate* Vol. 103, 1974 pp. 343–368
- MAGARI R.: Sur certe teorie non enumerabili, *Annali di Mat. pure ed applicate* Vol. 103, 1974 pp. 119–152

- MAGARI R.: The diagonalizable algebras (The algebraization of the theories which express Theor. II), *Boll Unione Mat Ital (IV Ser.)* Vol. 12, 1975 pp. 117-125
- MAGARI R.: Representation and duality theory for diagonalizable algebras (The algebraization of theories which express Theor; IV.), *Studia Logica* Vol. 34, 1975 pp. 305-313
- MAGARI R.: Problemi aperti sulle algebre diagonali, *Rend. Sem. Mat. Fis. Milano* Vol. 44, 1975 pp. 75-90
- MAGARI R.: Metodi algebrici in teoria della dimostrazione, *Boll. Un. Math. Ital.* Vol. 12, 1975 pp. 252-261
- MAGARI R.: On the autological character of diagonalizable algebras, *Studia Logica* Vol. 35, 1976 pp. 327-333
- MAGARI R.: Modal diagonalizable algebras, *Boll. Un. Math. Ital. B* Vol. 15, 1978 pp. 303-320
- MALYAUKENE L. K.: A constructive proof of the replaceability of the induction axiom in the free variable multiplicative arithmetic, *Litovsk. Mat. Sb.* Vol. 23, 1983 pp. 78-93
- MALYAUKENE L. K.: On the question of the relation between some weak inductions, *Mat. Logika Primen.* Vol. 3, 1983 pp. 34-37
- MANDERS K. L.: Computational complexity of decision problems in elementary number theory. In: *Model Theory of Algebra and Arithmetic (Lect. Notes Math. 834)*, Springer-Verlag, 1980 pp. 211-227
- MANDERS K. L., ADLEMAN L.: NP-complete decision problems for binary quadratics, *Journ. Comp. Sys. Sci.* Vol. 15, 1978 pp. 168-184
- MANEVITZ L. M.: Internal end-extensions of Peano arithmetic and a problem of Gaifman, *Journ. London Math. Soc., Ser 2* Vol. 13, 1976 pp. 80-82
- MANEVITZ L. M., STAVI J.: Δ_2^0 operators and alternating sentences in arithmetic, *Journ. Symb. Log.* Vol. 45, 1980 pp. 144-154
- MAR'TJANOV V. I.: Universal extended theories of integers, *Algebra i Logika* Vol. 16, 1977 pp. 588-602
- MARGENSTERN M.: Le théorème de Matiyassevitch et resultats connexes. In: *Model Theory and Arithmetic Comptes Rendus d'une Action Thématique Programmée du C. N. R. S. sur la Théorie des Modèles et l'Arithmétique*, Paris, France, 1979/80 (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 198-241
- MARKER D.: Degrees of models of true arithmetic. In: *Proceedings of the Herbrand Symposium Logic Colloquium '81 (Stud Logic Found Math 107)*, North Holland: Amsterdam, 1982 pp. 233-242
- MATYASEVIČ YU.: Arithmetic representations of powers. *Zap. naučn. sem. Leningr. otдел. Mat. in-ta V. A. Steklova* Vol. 8, 1968 pp. 159-165
- MATYASEVIČ YU.: Two reductions of Hilbert's tenth problem, *Zap. naučn. sem. Leningr. otдел. Mat. in-ta V. A. Steklova* Vol. 8, 1968 pp. 144-158
- MATYASEVIČ YU.: The connection between Hilbert's tenth problem and systems of equations between words and lengths, *Zap. naučn. sem. Leningr. otдел. Mat. in-ta V. A. Steklova* Vol. 8, 1968 pp. 132-144
- MATYASEVIČ YU.: The Diophantineness of enumerable sets, *Doklady Akad. Nauk SSSR* Vol. 191, 1970 pp. 279-282
- MATYASEVIČ YU.: Diophantine representation of enumerable predicates, *Izv. AN SSSR* Vol. 35, 1971 pp. 3-30
- MATYASEVIČ YU.: Diophantine representation of the set of prime numbers, *Doklady Akad. Nauk SSSR* Vol. 196, 1971 pp. 770-773
- MATYASEVIČ YU.: Diophantine representation of recursively enumerable predicates. In: *Actes du Congrès International des Mathématiciens 1970*, Gauthier Villars: Paris, 1971, pp. 235-238
- MATYASEVIČ YU.: Diophantine representation of recursively enumerable predicates. In: *Proceedings of the Second Scandinavian Logic Symposium*, North-Holland: Amsterdam, 1971, pp. 171-177

- MATIYASEVIČ YU.: Arithmetical representations of enumerable sets with a small number of quantifiers, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 32, 1972 pp. 77-84
- MATIYASEVIČ YU.: Diophantine sets, *Uspekhi Matem. Nauk.* Vol. 27, 1972 pp. 185-222
- MATIYASEVIČ YU.: On recursive unsolvability of Hilbert's tenth problem. In: *Logic, Methodology and Philosophy of Science IV*, North-Holland: Amsterdam, 1973, pp. 89-110
- MATIYASEVIČ YU.: The existence of non-effectivizable bounds in the theory of exponential Diophantine equations, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 40, 1974 pp. 77-93
- MATIYASEVIČ YU.: A new proof of the theorem on exponential Diophantine representation of recursively enumerable predicates, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 60, 1976 pp. 75-92
- MATIYASEVIČ YU.: Primes are non-negative values of a polynomial in 10 variables, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 68, 1977 pp. 62-82
- MATIYASEVIČ YU.: Algorithmic undecidability of exponential Diophantine equations in three unknowns. In: *Studies in the Theory of Algorithms and Mathematical Logic*, Moscow, 1979 pp. 69-78, 133
- MATIYASEVIČ YU., ROBINSON J.: Two universal three-quantifier representations of enumerable sets. In: *Theory of Algorithms and Mathematical Logic*, Comp. Center AN SSSR: Moscow, 1974, pp. 112-123, 216
- MATIYASEVIČ YU., ROBINSON J.: Reduction of an arbitrary Diophantine equation to one in 13 unknowns, *Acta Arithmetica* Vol. 27, 1975 pp. 521-553
- MCALOON K.: Applications alternées de théorèmes d'incomplétude et des théorèmes de complétude, *C. R. Acad. Sci, Paris (Ser A-B)* Vol. 280, 1975 pp. A849-A852
- MCALOON K.: Formules de Rosser pour ZF, *C. R. Acad. Sci.* Vol. 281, 1975 pp. A669-A672
- MCALOON K.: Consistency statements and number theories. In: *Colloq. int. de logique*, CNSR, Paris, 1977 pp. 199-207
- MCALOON K.: Completeness theorems, incompleteness theorems and models of arithmetic, *Trans. Amer. Math. Soc.* Vol. 239, 1978 pp. 253-277
- MCALOON K.: Diagonal methods and strong cuts in models of arithmetic. In: *Logic colloquium '77*, Proc, Wroclaw 1977, *Stud. Logic Found. Math.* Vol. 96, 1978 pp. 171-181
- MCALOON K.: Formes combinatoires du théorème d'incomplétude (d'après J. Paris et d'autres). In: *Seminaire Bourbaki* vol. 1977/1978 Exposes 507-524, Springer-Verlag, 1979, pp. 263-276
- MCALOON K.: Progressions transfinies de theories axiomatiques, formes combinatoires du théorème d'incomplétude et fonctions recursives a croissance rapide. In: *Models of arithmetic* (Sem, Univ. Paris VII, Paris, 1977), *Soc. Math. France, Paris*, 1980 pp. 41-58
- MCALOON K.: Les rapports entre la méthode des indicatrices et la méthode de Gödel pour obtenir des resultats d'indépendance. In: *Models of arithmetic* (Sem. Univ. Paris VII, Paris, 1977), *Soc. Math. France, Paris*, 1980 pp. 31-39
- MCALOON K.: On the complexity of models of arithmetic. *Journ. Symb. Log.* Vol. 47, 1982 pp. 403-415
- MCALOON K.: Paris-Harrington incompleteness and progressions of theories. In: *Recursion theory* (Ithaca, N. Y, 1982), *Amer. Math. Soc, Providence, RI*, 1985 pp. 447-460
- MCALOON K.: Models of arithmetic and complexity theory. In: *Studies in Complexity Theory*, Wiley, 1986 pp. 119-221
- MCALOON K. (ED): *Modèles de l'arithmétique*, Soc Math France: Paris, 1980, 155 p.
- MCALOON K., RESSAYRE J. P.: Les methodes de Kirby-Paris et la theorie des ensembles. In: *Model Theory and Arithmetic. Comptes Rendus d'une Action Thématique Programmée du CNRS sur la théorie des Modèles et l'Arithmétique.* (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 154-184

- MENDELSON E.: On non-standard models for number theory. In: *Essays on the Foundations of Mathematics: Dedicated to A. A. Fraenkel on His 70th Anniversary*, Magnes Pr: Jerusalem, 1961 pp. 259-268
- MENDELSON E.: Introduction to mathematical logic. D. van Nostrand Comp. Inc., Princeton 1964, 300 pp.
- MESCHKOWSKI H.: *Problemgeschichte der Mathematik I-III* Bibl. Inst. Mannheim, 1981
- MEYER A. R.: Weak monadic second order theory of successor is not elementary-recursive. In: *Logic Colloquium, Symposium on Logic Held at Boston, 1972-73* (Lect. Notes Math. 453), Springer-Verlag, 1973 pp. 132-154
- MICHEL P.: Borne supérieure de la complexité de N muni de la relation et divisibilité. In: *Model Theory and Arithmetic* (Lect. Notes Math. 890), Springer-Verlag 1981 pp. 242
- MIJAJLOVIĆ Z.: Submodels and definable points in models of Peano arithmetics, *Notre Dame Journ. Formal Log.* Vol. 24, 1983 pp. 417-425
- MILLER C. F.: Some connections between Hilbert's 10th problem and the theory of groups. In: *Word Problems*, North-Holland: Amsterdam, 1973
- MILLS G.: Extensions of models of Peano arithmetic. Dissertation, Berkeley, 1977
- MILLS G.: A model of Peano arithmetic with no elementary end extension, *Journ. Symb. Log.* Vol. 43, 1978 pp. 563-567
- MILLS G.: Substructure lattices of models of arithmetic, *Ann. Math. Logic* Vol. 16, 1979 pp. 145-180
- MILLS G.: A tree analysis of unprovable combinatorial statements. In: *Model theory of algebra and arithmetic* (Proc. Conf, Karpacz, 1979) (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 248-311
- MILLS G., PARIS J. B.: Closure properties of countable nonstandard integers, *Fund. Math.* Vol. 103, 1979 pp. 205-215
- MILLS G., PARIS J. B.: Regularity in models of arithmetic, *Journ. Symb. Log.* Vol. 49, 1984 pp. 272-280
- MINC G. E.: Priloženje teoremy Herbranda. In: *Matematičeskaja teorija logičeskogo vyvoda*, Nauka: Moscow, 1967 pp. 311-350
- MINC G. E.: Exact estimate of the provability of transfinite induction in initial parts of arithmetic, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 20, 1971 pp. 134-144
- MINC G. E.: Quantifier-free and one-quantifier systems, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 20, 1971 pp. 115-133
- MINC G. E.: Finite investigations of infinite derivations, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 49, 1975 pp. 67-122
- MINC G. E.: What can be done in PRA, *Zap. naučn. sem. Leningr. otdel. Mat. in-ta V. A. Steklova* Vol. 60, 1976 pp. 93-102
- MISERCQUE D.: Sur le treillis des $\forall_1$ -formules fermées de l'arithmétique de Peano, *Compte Rendus, Serie A* Vol. 290, 1980 pp. 571-573
- MISERCQUE D.: Solutions de deux problèmes posés par H. Simmons. In: *Proceedings of the Model Theory Meeting* (Univ. Brussels, Brussels/Univ. Mons, Mons, 1980), *Bull. Soc. Math. Belg. Ser. B* Vol. 33, 1981 pp. 65-72
- MISERCQUE D.: Answer to a problem of D. Guaspari. In: *Proceedings of the International Conference "Open Days in Model Theory and Set Theory"* (Jadwisin), 1981 pp. 181-184
- MISERCQUE D.: The nonhomogeneity of the E -tree - answer to a problem raised by D. Jensen and A. Ehrenfeucht, *Proc. Amer. Math. Soc.* Vol. 84, 1982 pp. 573-575
- MIYATAKE T.: On the length of proofs in formal systems, *Tsukuba Journ. Math.* Vol. 4, 1980 pp. 115-125
- MIYATAKE T.: On the length of proofs in a formal system of recursive arithmetic. In: *Logic Symposia Hakone 1979, 1980* (Lect. Notes Math. 891), Springer-Verlag, 1981 pp. 81-108
- MLČEK J.: A representation of models of Peano arithmetic, *Comm. Math. Univ. Carol.* Vol. 14, 1973 pp. 553-558

- MLČEK J.: Twin prime problem in an arithmetic without induction, *Comm. Math. Univ. Carol.* Vol. 17, 1976 pp. 543–555
- MLČEK J.: End-extensions of countable structures and the induction schema, *Comm. Math. Univ. Carol.* Vol. 19, 1978 pp. 291–308
- MLČEK J.: A note on confinal extensions and segments, *Comm. Math. Univ. Carol.* Vol. 19, 1978 pp. 727–742
- MONTAGNA F.: For every n , the n -freely generated algebra is not functionally free in the equational class of diagonalizable algebras (the algebraization of theories which express Theor. V), *Studia Logica* Vol. 34, 1975 pp. 315–319
- MONTAGNA F.: On the algebraization of Feferman's predicate, *Studia Logica* Vol. 37, 1978 pp. 221–236
- MONTAGNA F.: On the algebraization of Feferman's predicate, *Studia Logica* Vol. 37, 1978 pp. 221–236
- MONTAGNA F.: On the formulas of Peano arithmetic which are provably closed under modus ponens, *Boll. Unione Mat. Ital.* Vol. 16, 1979 pp. 196–211
- MONTAGNA F.: On the diagonalizable algebra of Peano arithmetic, *Boll. Un. Math. Ital.* Vol. 16-B, 1979 pp. 795–812
- MONTAGNA F.: On the diagonalizable algebra of Peano arithmetic, *Bolletino U. M. I.* Vol. 16-B, 1979 pp. 795–812
- MONTAGNA F.: Interpretations of the first-order theory of diagonalizable algebras in Peano arithmetic, *Studia Logica* Vol. 39, 1980 pp. 347–354
- MONTAGNA F.: The predicate modal logic of provability. *Notre Dame Journ. Formal Log.* Vol. 25, 1984 pp. 179–192
- MONTAGNA F.: A completeness result for fixed-point algebras, *Zeitschr. Math. Log. Grundl. Math.* Vol. 30, 1984 pp. 525–532
- MONTAGNA F.: Provability in finite subtheories of PA and relative interpretability: a modal investigation, *Journ. Symb. Log.* Vol. 52, 1987 pp. 494–511
- MONTAGNA F.: Iterated extensional Rosser's points and hyperhyperdiagonalizable algebras, *Zeitschr. Math. Log. Grundl. Math.* Vol. 33, 1987 pp. 293–303
- MONTAGNA F., SOMMARUGE G.: Rosser and Mostowski sentences, *Arch. Math. Log.* Vol. 27, 1988 pp. 115–133
- MONTAGNA F., SORBI A.: Universal recursion theoretic properties of r. e. preordered structures, *Journ. Symb. Log.* Vol. 50, 1985 pp. 397–406
- MONTAGUE R.: Non-finite axiomatizability. In: *Summaries of Talks Presented at the summer Institute for Symbolic Logic, Institute for Defense Analyses, Comm. Res. Division*, 1957 pp. 256–259
- MONTAGUE R.: Two theorems on relative interpretability. In: *Summer Institut of Symbolic Logic, Cornell Univ.*, 1957 pp. 263–264
- MONTAGUE R.: Semantical closure and non-finite axiomatizability I. In: *Infinitistic Methods*, Pergamon Press: London, 1961 pp. 45–69
- MONTAGUE R.: Theories incomparable with respect to relative interpretability, *Journ. Symb. Log.* Vol. 27, 1962 pp. 195–211
- MONTAGUE R.: Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philos Fenn* Vol. 16, 1963 pp. 153–167
- MONTAGUE R.: Interpretability in terms of models, *Indagationes Mathematicae* Vol. 27, 1965 pp. 467–476
- MONTAGUE R., TARSKI A.: Independent recursive axiomatizability. In: *Summer Institut of Symbolic Logic, Cornell Univ.*, 1957 pp. 270
- MORGENSTERN C.: On generalized quantifiers in arithmetic, *Journ. Symb. Log.* Vol. 47, 1982 pp. 187–190
- MORTENSEN C.: Inconsistent nonstandard arithmetic, *Journ. Symb. Log.* Vol. 52, 1987 pp. 512–518
- MORTENSEN C.: Inconsistent number systems, *Notre Dame Journ. Formal Log.* Vol. 29, 1988 pp. 45–60

- MOSTOWSKI A.: Bemerkungen zum Begriff der inhaltlichen Widerspruchsfreiheit, *Journ. Symb. Log.* Vol. 4, 1939 pp. 113–114
- MOSTOWSKI A.: On definable sets of positive integers, *Fund. Math.* Vol. 34, 1947 pp. 81–112
- MOSTOWSKI A.: An undecidable arithmetical statement, *Fund. Math.* Vol. 36, 1949 pp. 143–164
- MOSTOWSKI A.: On models of axiomatic systems, *Fund. Math.* Vol. 39, 1952 pp. 133–158
- MOSTOWSKI A.: Sentences undecidable in formalized arithmetic. An exposition of the theory of Kurt Gödel, North Holland: Amsterdam, 1952, 117 p.
- MOSTOWSKI A.: On direct products of theories, *Journ. Symb. Log.* Vol. 17, 1952 pp. 1–31
- MOSTOWSKI A.: On a system of axioms which has no recursively enumerable model, *Fund. Math.* Vol. 40, 1953 pp. 56–61
- MOSTOWSKI A.: On recursive models of formalised arithmetic, *Bull. Acad. Polon. Sci.* Vol. 5, 1957 pp. 705–710
- MOSTOWSKI A.: On various degrees of constructivism. In: Heyting (ed.), *Constructivism in mathematics*, North-Holland Publ. Comp. 1959, pp. 178–194
- MOSTOWSKI A.: A generalization of the incompleteness theorem, *Fund. Math.* Vol. 49, 1961 pp. 205–232
- MOSTOWSKI A.: Representability of sets in formal systems. In: *Recursive Function Theory*, Amer. Math. Soc.: Providence, 1962 pp. 29–48
- MOSTOWSKI A., ROBINSON R. M., TARSKI A.: Undecidability and essential undecidability in arithmetic. In: *Undecidable Theories (Stud Logic Found Math)*, North Holland: Amsterdam, 1953 pp. 39–74
- MÜLLER G. H.: Über die unendliche Induktion. In: *Infinitistic Methods. Proceedings of the Symposium on Foundations of Mathematics*, PWN: Warsaw, Pergamon Pr: Oxford, 1961 pp. 75–95
- MÜLLER G. H.: Nicht-Standardmodelle der Zahlentheorie. *Zeitschr. Math. Log. Grundl. Math.* Vol. 77, 1961 pp. 414–438
- MÜLLER G. H. (ED.): Ω -bibliography of mathematical logic, vols I–VI, Springer-Verlag, 1987
- MUNDICI D.: Natural limitations of decision procedures for arithmetic with bounded quantifiers, *Arch. Math. Log.* Vol. 23, 1983 pp. 37–54
- MURAWSKI R.: Expandability of models for elementary arithmetic, *Komunikaty i Rozprawy Inst: Poznan*, 1976, 26 p.
- MURAWSKI R.: On expandability of models of Peano arithmetic. I, II, *Studia Logica* Vol. 35, 1976 pp. 409–431
- MURAWSKI R.: On expandability of models of Peano arithmetic. III, *Studia Logica* Vol. 36, 1977 pp. 181–188
- MURAWSKI R.: Indicator, satisfaction classes and expandability, *Komunikaty i Rozprawy Inst: Poznan*, 1978, 17 p.
- MURAWSKI R.: Indicators and the structure of expansions, *Komunikaty i Rozprawy Inst: Poznan*, 1979, 18 p.
- MURAWSKI R.: Some remarks on the structure of expansions, *Zeitschr. Math. Log. Grundl. Math.* Vol. 26, 1980 pp. 537–546
- MURAWSKI R.: Incompleteness of Σ_n^O definable theories via indicators, *Funct. Approx. Comment. Math.* Vol. 11, 1981 pp. 57–63
- MURAWSKI R.: A note on inner interpretations of models of Peano arithmetic, *Rep. Math. Logic* Vol. 13, 1981 pp. 53–57
- MURAWSKI R.: Some more remarks on expandability of initial segments. In: *Proceedings of the International Conference "Open Days in Model Theory and Set Theory"*, 1981 pp. 221–230
- MURAWSKI R.: A simple remark on satisfaction classes, indiscernibles and recursive saturation, *Fct Approximatio, Comment Math, Poznan* Vol. 11, 1981 pp. 149–151

- MURAWSKI R.: A contribution to nonstandard teratology. In: Models and sets, Proc. Logic Colloq, Aachen 1983, Part I, (Lect. Notes Math. 1103), Springer-Verlag, 1984 pp. 379–388
- MURAWSKI R.: Expandibility of models of arithmetic. In: Frege conference, Proc. Int. Conf, Schwerin / Ger. 1984, Math. Res. 20, 1984 pp. 87–93
- MURAWSKI R.: Trace expansions of initial segments, Zeitschr. Math. Log. Grundle. Math. Vol. 30, 1984 pp. 471–476
- MURAWSKI R.: Giuseppe Peano-pioneer and promoter of symbolic logic. In: Komunikaty i Rozprawy, Univ im. A. Mickiewicza, Inst. Matematyki, 1985 pp. 21
- MURAWSKI R.: On expansions of substructures. (abstract), Journ. Symb. Log. Vol. 51, 1986 pp. 493
- MURAWSKI R.: Some more remarks on expandability of initial segments, Zeitschr. Math. Log. Grundle. Math. Vol. 32, 1986 pp. 445–450
- MURAWSKI R.: Pointwise definable substructures of models of Peano arithmetic, Notre Dame Journ. Formal Log. Vol. 29, 1988 pp. 295–308
- MURAWSKI R.: Definable sets and expansions of models of Peano arithmetic, Arch. Math. Log. Vol. 27, 1988 pp. 21–33
- MYCIELSKI J.: The definition of arithmetic operations in the Ackermann model, Algebra i Logika Vol. 3/5-6, 1964 pp. 64–65
- MYCIELSKI J.: A lattice of interpretability types of theories, Journ. Symb. Log. Vol. 42, 1977 pp. 297–305
- MYCIELSKI J.: Locally finite theories, Journ. Symb. Log. Vol. 51, 1986 pp. 59–62
- MYCIELSKI J.: Finitistic consistency proofs, Univ. of Colorado, preprint, 1982
- MYCIELSKI J.: The meaning of pure mathematics, Journ. Phil. Log., Vol. 18, 1989, pp. 315–320
- MYCIELSKI J., PUDLÁK P., STERN A. S.: A lattice of chapters of mathematics, Memoirs Amer. Math. Soc. Vol. 84, 1990, 70 pp.
- MYHILL J. R.: A system which can define its own truth. Fund. Math. Vol. 37, 1950 pp. 190–192
- MYHILL J. R.: An absolutely independent set of Σ_1^0 sentences, Zeitschr. Math. Log. Grundle. Math. Vol. 18, 1972 pp. 107–109
- NADEL M. E.: On a problem of MacDowell and Specker, Journ. Symb. Log. Vol. 45, 1980 pp. 612–622
- NADEL M. E.: The completeness of Peano multiplication. Israel Journ. Math. Vol. 39, 1981 pp. 225–233
- NEGRI M.: An application of recursive saturation, Boll. Un. Mat. Ital. A (6) Vol. 3, 1984 pp. 449–451
- NEGRI M.: On the non finite axiomatizability of P. In: Atti degli incontri di logica matematica, Vol. 2, Siena / Italia 1983/84, 1985 pp. 621–626
- NELSON E.: Predicative arithmetic. Mathematical Notes, Princeton Univ. Press, 1986, 189 p.
- NELSON G. C.: A further restricted ω -rule, Colloq. Math. Vol. 23, 1971 pp. 1–3
- NEPOMNJAŠČIJ V. A.: Rudimentary predicates and Turing calculations, Dokl. Akad. Nauk SSSR, Tom 195, 1970; Soviet Math. Dokl. Vol. 11, No. 6, 1970 pp. 1462–1465
- O'DONNELL M.: A programming language theorem which is independent of Peano arithmetic. In: A. C. M. Symp, 1979 pp. 176–188
- OKADA M.: A simple relationship between Buchholz's new system of ordinal notations and Takeuti's system of ordinal diagrams, Journ. Symb. Log. Vol. 52, 1987 pp. 577–581
- ONO H.: Reflection principles in fragments of Peano arithmetic, Zeitschr. Math. Log. Grundle. Math. Vol. 33, 1987 pp. 317–333
- ONO H., KADOTA N.: Provably recursive functions in fragments of Peano arithmetic. (abstract), Journ. Symb. Log. Vol. 51, 1986 pp. 494

- OREVKOV V. P.: Lower bounds to derivations in arithmetic via complexity of the terms involved in derivations, Russian, *Doklady Akad. Nauk SSSR*, Vol. 294, 1987, pp. 784–787
- OREVKOV V. P.: Reconstruction of a proof by its analysis, Russian, *Doklady Akad. Nauk SSSR* Vol. 293, 1987 pp. 313–316
- OREY S.: On ω -consistency and related properties, *Journ. Symb. Log.* Vol. 21, 1956 pp. 246–252
- OREY S.: Relative interpretations, *Journ. Symb. Log.* Vol. 24, 1959 pp. 281–282
- OREY S.: Relative interpretations, *Zeitschr. Math. Log. Grundle. Math.* Vol. 7, 1961 pp. 146–153
- PABION J. F.: L'axiomatisation de la syntaxe et le second théorème de Gödel, *Publ. Dep. Math. Lyon* Vol. 11, 1974 pp. 27–87
- PABION J. F.: Saturation en arithmétique et en analyse, *Bull. Soc. Math. Belg. Ser B* Vol. 33, 1981 pp. 73–82
- PABION J. F.: Saturated models of Peano arithmetic, *Journ. Symb. Log.* Vol. 47, 1982 pp. 625–637
- PABION J. F., RICHARD D.: Synonymy and re-interpretation for some sublanguages of Peano arithmetic. In: *Proceedings of the International Conference "Open Days in Model Theory and Set Theory"*, 1981 pp. 231–236
- PALÚCH S.: The lattices of numerations of theories containing Peano's arithmetic, *Comm. Math. Univ. Carol.* Vol. 14, 1973 pp. 339–359
- PAOLA DI R. A.: On sets represented by the same formula in distinct consistent axiomatizable Rosser theories, *Pac. Journ. Math.* Vol. 18, 1966 pp. 455–456
- PAOLA DI R. A.: Pseudo-complements and ordinal logics based on consistency statements, *Journ. Symb. Log.* Vol. 31, 1966 pp. 359–364
- PAOLA DI R. A.: Some properties of pseudo-complements of recursively enumerable sets, *Trans. Amer. Math. Soc.* Vol. 121, 1966 pp. 296–308
- PAOLA DI R. A.: Some theorems on extensions of arithmetic, *Journ. Symb. Log.* Vol. 32, 1967 pp. 180–189
- PAOLA DI R. A.: A note on diminishing the undecidable region of a recursively enumerable set, *Duke Math. Journ.* Vol. 35, 1968 pp. 399–406
- PAOLA DI R. A.: A theorem on shortening the length of proof in formal systems of arithmetic, *Journ. Symb. Log.* Vol. 40, 1975 pp. 398–400
- PAOLA DI R. A.: A uniformly, extremely nonextensional formula of arithmetic with many undecidable fixed points in many theories, *Proc. Amer. Math. Soc.* Vol. 92, 1984 pp. 291–297
- PAPPINGHAUS P.: A version of the Σ_1 principle for CFA provable in PRA, *Arch. Math. Log.* Vol. 20, 1980 pp. 27–40
- PARIKH R.: Existence and feasibility in arithmetic, *Journ. Symb. Log.* Vol. 36, 1971 pp. 494–508
- PARIKH R.: Some results on the length of proofs, *Trans. Amer. Math. Soc.* Vol. 177, 1973 pp. 29–36
- PARIS J. B.: On models of arithmetic. In: *Conference on Mathematical Logic – London '70* (Lect. Notes Math. 255), Springer-Verlag, 1972 pp. 251–280
- PARIS J. B.: Models of arithmetic and the 1-3-1 lattice, *Fund. Math.* Vol. 95, 1977 pp. 195–199
- PARIS J. B.: Note on an induction axiom, *Journ. Symb. Log.* Vol. 43, 1978 pp. 113–117
- PARIS J. B.: Some independence results for Peano arithmetic, *Journ. Symb. Log.* Vol. 43, 1978 pp. 725–731
- PARIS J. B.: A hierarchy of cuts in models of arithmetic. In: *Model theory of algebra and arithmetic*, (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 312–337
- PARIS J. B.: Some conservation results for fragments of arithmetic. In: *Model theory of algebra and arithmetic*, (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 251–262

- PARIS J. B.: O struktuře modelu omezené E_1 indukce (On the structure of models of bounded E_1 -induction), *Časopis pěst. mat.* Vol. 109, 1984 pp. 372–379
- PARIS J. B., DIMITRACOPOULOS C.: Truth definitions for Δ_0 formulae. In: *Logic and algorithmic*, *L'enseignement Mathématique* No 30, Geneve, 1982 pp. 318–329
- PARIS J. B., DIMITRACOPOULOS C.: A note on the undefinability of cuts, *Journ. Symb. Log.* Vol. 48, 1983 pp. 564–569
- PARIS J. B., HARRINGTON L.: A mathematical incompleteness in Peano arithmetic, *Handbook of Mathematical Logic*, North-Holland P. C, 1977 pp. 1133–1142
- PARIS J. B., KIRBY L. A. S.: Σ_n -collection schema in arithmetic. In: *Logic Colloquium 1977* (Studies in Logic and the Foundations of Mathematic 96), North-Holland P. C, 1978 pp. 199–209
- PARIS J. B., MILLS G.: Closure properties of countable non-standard integers, *Fund. Math.* Vol. 103, 1979 pp. 205–215
- PARIS J. B., WILKIE A. J.: Δ_0 sets and induction. In: *Proceedings of the Jadwisin Logic Conference*, Poland, Leeds University Press, 1981 pp. 237–248
- PARIS J. B., WILKIE A. J.: Models of arithmetic and the rudimentary sets, *Bull Soc Math Belg*, Ser B Vol. 33, 1981 pp. 157–169
- PARIS J. B., WILKIE A. J.: Some results on bounded induction. In: *Proceedings of the 2nd Easter Conference on Model Theory*, Humboldt-Univ. Berlin: Berlin, 1984 pp. 223–228
- PARIS J. B., WILKIE A. J.: Counting problems in bounded arithmetic. In: *Methods in mathematical logic*, Proc. 6th Latin Amer. Symp, Caracas / Venez. 1983, (Lect. Notes Math. 1130), Springer-Verlag, 1985 pp. 317–340
- PARIS J. B., WILKIE A. J.: Counting Δ_0 sets, *Fund. Math.* Vol. 127, 1987 pp. 67–76
- PARIS J. B., WILKIE A. J.: Δ_0 approximations to the counting functionals, preprint
- PARIS J. B., WILKIE A. J.: On the existence of end extensions of models of bounded induction. In: *Logic, Methodology and Philosophy of Science VIII*, 1983, North-Holland, to appear
- PARIS J. B., WILKIE A. J., WOODS A. R.: Provability of the Δ_0 -PHP and the existence of infinitely many primes, *Journ. Symb. Log.* Vol. 53, 1988 pp. 1235–1244
- PARLAMENTO F.: Binumerability in a sequence of theories, *Rend. Sem. Mat. Univ. Padova* Vol. 65, 1981 pp. 9–12
- PARLAMENTO F.: PRA provability of Schmerl's fine structure theorem, *Rend. Sem. Mat., Torino* Vol. 42, 1984 pp. 87–106
- PARSONS C.: On number-theoretic choice schema and its relation to induction. In: *Intuitionism and Proof Theory*, North Holland: Amsterdam, 1970 pp. 459–473
- PARSONS C.: On n -quantifier induction, *Journ. Symb. Log.* Vol. 37, 1972 pp. 466–482
- PEANO G.: *Arithmetices principia nuovo methodo exposita* (The principles of arithmetic, presented by a new method), Bocca & Clausen, 1889, 20 p.
- PEANO G.: *Formulario matematico*, Torino, 1908, 463 pp.
- PENZIN Y. G.: The problem of twin primes in formal arithmetic, *Mat. Zametki* Vol. 26, 1979 pp. 505–511
- PENZIN Y. G.: Twins problem in formal arithmetic, *Math. Notes* Vol. 26, 1980 pp. 743–746
- PHILLIPS R. G.: On the structure of non-standard models of arithmetic, *Proc. Amer. Math. Soc.* Vol. 27, 1971 pp. 359–363
- PHILLIPS R. G.: Addition in nonstandard models of arithmetic, *Journ. Symb. Log.* Vol. 37, 1972 pp. 483–486
- PHILLIPS R. G.: Omitting types in arithmetic and conservative extensions. In: *Victoria Symposium on Non-Standard Analysis* (Lect. Notes Math. 369), Springer-Verlag, 1974 pp. 195–202
- PHILLIPS R. G.: A minimal extension that is not conservative, *Mich. Math. Journ.* Vol. 21, 1974 pp. 27–32
- PILLAY A.: Partition properties and definable types in Peano arithmetic. In: *Model theory and arithmetic* (Paris, 1979–1980) (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 263–269

- PILLAY A.: Models of Peano arithmetic (a survey of basic results). In: *Model Theory and Arithmetic. Comptes Rendus d'une Action Thematique Programmee du CNRS sur la Theorie des Modeles et l'Arithmetique* (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 1-12
- PILLAY A.: Cuts in models of arithmetic. In: *Model theory and Arithmetic. Comptes Rendus d'une Action Thematique Programmee du CNRS sur la Theorie des Modeles et l'Arithmetique* (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 13-20
- PLISKO V. E.: The nonarithmeticity of the class of realizable predicate formulas, *Math. USSR Vol. Izv.* 11, 1977 pp. 453-471
- POHLERS W.: An upper bound for provability of transfinite induction in systems with n -times iterated inductive definitions. In: *Proof Theory Symposium, Kiel 1974* (Lect. Notes. Math. 500), Springer-Verlag, 1975 pp. 271-289
- POHLERS W.: Ordinal notations based on a hierarchy of inaccessible cardinals, *Annals Pure Appl. Logic* Vol. 33, 1987 pp. 157-180
- POHLERS W.: *Proof Theory. An Introduction.* Springer-Verlag 1989, Lect. Notes Math. 1407, 213 p.
- POIZAT B.: *Cours de théorie des modèles*, Nur al-Manteq, wal-Marifah, Villenobanne, 1985 pp. 142-207
- POTTHOFF K.: *Untersuchungen über Nichtstandardmodelle.* In: *Dissertation, Hannover*, 1967
- POTTHOFF K.: Über Nichtstandardmodelle der Arithmetik und der rationalen Zahlen, *Zeitschr. Math. Log. Grundl. Math.* Vol. 15, 1969 pp. 223-236
- POTTHOFF K.: Ideale in Nichtstandardmodellen der ganzen Zahlen, *Zeitschr. Math. Log. Grundl. Math.* Vol. 16, 1970 pp. 321-326
- POTTHOFF K.: Über Erweiterungen von Nichtstandardmodellen der Arithmetik und anderer Strukturen. In: *Diss, Habil. etc.*, 1972
- POTTHOFF K.: Ordnungseigenschaften von Nichtstandardmodellen. In: *Theory of Sets and Topology. In Honour of Felix Hausdorff (1868-1942)*, Dt. Verlag Wiss.: Berlin, 1972 pp. 403-426
- POTTHOFF K.: A simple tree lemma and its application to a counterexample of Phillips, *Arch. Math. Log.* Vol. 18, 1976 pp. 67-71
- POTTHOFF K.: Orderings of type of countable arithmetic, *Zeitschr. Math. Log. Grundl. Math.* Vol. 24, 1978 pp. 97-108
- PRAWITZ D.: Ideas and results in proof theory. In: *Proceedings of the Second Scandinavian Logic Symposium, North Holland: Amsterdam* pp. 235-307
- PRAWITZ D.: The philosophical position of proof theory. In: *Contemporary Philosophy in Scandinavia*, The Johns Hopkins Press, 1972 pp. 123-134
- PRESBURGER M.: Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In: *Sprawozdanie z 1 Kongresu Matematyków Krajow Slowiańskich*, Ksiaznica Atlas: Warsaw 1930 pp. 92-101, 395
- PUDLÁK P.: A definition of exponentiation by a bounded arithmetical formula, *Comm. Math. Univ. Carol.* Vol. 24, 1983 No. 4 pp. 667-671
- PUDLÁK P.: Some prime elements in the lattice of interpretability types, *Trans. Amer. Math. Soc.* Vol. 280, 1983 No. 1 pp. 255-275
- PUDLÁK P.: Cuts, consistency statements and interpretations, *Journ. Symb. Logic* Vol. 50, 1985 pp. 423-441
- PUDLÁK P.: On the length of proofs of finitistic consistency statements in first order theories. In: *Logic Colloquium 84, North Holland P. C.*, 1986 pp. 165-196
- PUDLÁK P.: Improved bounds to the length of proofs of finitistic consistency statements, *Contemporary mathematics* Vol. 65, 1987 pp. 309-331
- PUDLÁK P.: A note on bounded arithmetic, *Fundamenta Mathematicae* 136, 1990 pp. 85-89

- PUDLÁK P.: Another combinatorial principle independent of Peano's axioms, (unpublished text)
- PUDLÁK P.: Some relations between subsystems of arithmetic and the complexity of computations. In: *Logic from Computer Science*, ed. Y. N. Moschovakis, Springer-Verlag 1992, pp. 499–519
- PUDLÁK P.: Ramsey's theorem in bounded arithmetic. In: *Computer Science Logic*, eds. Börger et al., Springer Lecture Notes in Comp. Sci. 553, 1992, pp. 308–317
- PUDLÁK P. SOCHOR A.: Models of the Alternative Set Theory, *Journ. Sym. Log.* Vol. 49, 1984 No. 2 pp. 570–585
- PURITZ C. W.: Ultrafilters and standard functions in non-standard arithmetic, *Proc London Math Soc*, Ser 3 Vol. 22, 1971 pp. 705–733
- PUTNAM H.: An unsolvable problem in number theory, *Journ. Symb. Log.* Vol. 25, 1960 pp. 220–232
- PUTNAM H.: Trial and error predicates and the solution to a problem of Mostowski, *Journ. Symb. Log.* Vol. 30, 1965 pp. 49–57
- PUTNAM H., SMULLYAN R.: Exact separation of recursively enumerable sets within theories, *Proc. Amer. Math. Soc.* Vol. 11, 1960 pp. 574–577
- QUINSEY J.: Some problems in logic. Dissertation, Oxford, 1980
- QUINSEY J.: Sets of Σ_k -conservative sentences are Π_2 -complete, *Journ. Symb. Log.* Vol. 46, 1981 pp. 442
- RABIN M. O.: Non-standard models and independence of the induction axiom. In: *Essays on the Foundations of Mathematics: Dedicated to A. A. Freinkel on his 70th anniversary*, North Holland, Amsterdam, 1961 pp. 287–299
- RABIN M. O.: Diophantine equations and non-standard models of arithmetic. In: *Proceedings of the 1st International Congress for Logic, Methodology and Philosophy of Science*, Stanford Univ Pr: Stanford, 1962 pp. 151–158 (Russian translation in *Matematicheskaja logika i ee primeneniya*, Mir 1965, 176–184)
- RACKOFF C.: On the complexity of theories of weak direct powers, *Journ. Symb. Log.* Vol. 41, 1976 pp. 561–573
- RAISONNIER J.: Ensembles mesurables et consistance de ZF, *C. R. Acad. Sci. Paris Ser. I Math.* Vol. 297, 1983 pp. 217–219
- RATAJCZYK Z.: Satisfaction classes and combinatorial sentences independent from PA, *Zeitschr. Math. Log. Grundle. Math.* Vol. 28, 1982 pp. 149–165
- RATAJCZYK Z.: A combinatorial analysis of functions provably recursive in $I\Sigma_n$, *Fund. Math.* vol 130, 1988, 191–213
- RATAJCZYK Z.: Functions provably total in $I\Sigma_n$, *Fund. Math.* Vol. 132, 1989, pp. 81–95
- RAZBOROV A. A.: An equivalence between second order bounded domain bounded arithmetic and first order bounded arithmetic. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 247–277
- REMMEL J. B.: Graph colorings and recursively bounded Π_1^0 -classes, *Annals Pure Appl. Logic* Vol. 32, 1986 pp. 185–193
- RESSAYRE J. P.: Types remarquables et extensions de modèles dans l'arithmétique de Peano II (Avec un appendice de M. A. Dickmann), *Astérisque* Vol. 73, 1980 pp. 119–155
- RESSAYRE J. P.: Modèles isomorphes à leurs propres segments initiaux en théorie des ensembles, *C. R. Acad. Sci, Paris, Ser 1* Vol. 296, 1983 pp. 569–572
- RESSAYRE J. P.: Non standard universes with strong embeddings, and their finite approximations. In: *Logic and Combinatorics, Proceedings of the Amer. Math. Soc. – IMS-SIAM Joint Summer Research Conference*, Amer. Math. Soc.: Providence, 1987 pp. 333–358
- RESSAYRE J. P.: Integer parts of real closed exponential fields (extended abstract). In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 278–288

- RESSAYRE J. P., WILKIE A. J.: Modèles non-standard en arithmétique et théorie des ensembles, Publ. Math. Univ. Paris VII, 1986, 147 p.
- RICHARD D.: On external properties of nonstandard models of arithmetic, Publ. Dep. Math. Vol. 14, 1977 pp. 57-75
- RICHARD D.: Saturation des modèles de Peano, C. R. Acad. Sci. Paris Sér. A-B Vol. 290, 1980 pp. A351-A353
- RICHARD D.: De la structure additive à la saturation des modèles de Peano et à une classification des sous-langages de l'arithmétique. In: Model theory and arithmetic (Paris, 1979-1980) (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 270-296
- RICHARD D.: Les relations arithmétiques sur les entiers primaires sont définissables au premier ordre par successeur et comprimarite, C. R. Acad. Sci. Paris, Ser. I Vol. 299, 1984 pp. 795-798
- RICHARD D.: The arithmetics as theories of two orders. In: Orders: description and roles, Proc. Conf. Ordered sets appl, l'Arbresle / France 1982, Ann. Discrete Math. 23, 1984 pp. 287-311
- RICHARDSON D.: Some undecidable problems involving elementary functions of a real variable, Journ. Symb. Log. Vol. 33, 1968 pp. 514-520
- RICHARDSON D.: Solution of the identity problem for integral exponential functions, Zeitschr. Math. Log. Grundle. Math. Vol. 15, 1969 pp. 333-340
- RICHARDSON D.: Sets of theorems with short proofs, Journ. Symb. Log. Vol. 39, 1974 pp. 235-242
- RIEGER L.: Sur le problème des nombres naturels. In: Infinitistic Methods. Proceedings of the Symposium on Foundations of Mathematics, Pergamon Press: Oxford, 1961 pp. 225-233
- RIIS S.: Making infinite structures finite in models of second order bounded arithmetic. In: Arithmetic, Proof Theory and Computational Complexity, eds. P. Clote and J. Krajčček, Oxford Univ. Press 1992, pp. 289-319
- RITCHIE R. W., YOUNG P. R.: Strong representability of partial functions in arithmetic theories, Information Sci. Vol. 1, 1968 pp. 189-204
- RITTER W. E.: Representability of partial recursive functions in formal theories, Proc. Amer. Math. Soc. Vol. 18, 1967 pp. 647-651
- ROBBIN J. W.: Subrecursive hierarchies. Ph. D. thesis, Princeton 1965.
- ROBINSON A.: Model theory and non-standard arithmetic. In: Infinitistic Methods. Proceedings of the Symposium on Foundations of Mathematics, PWN: Warsaw, 1961 pp. 265-302
- ROBINSON A.: On languages which are based on nonstandard arithmetic, Nagoya Math. Journ. Vol. 22, 1963 pp. 83-117
- ROBINSON A.: Infinite forcing in model theory. In: Proceedings of the Second Scandinavian Logic Symposium, North-Holland: Amsterdam, 1971, pp. 317-340
- ROBINSON A.: Nonstandard arithmetic and generic arithmetic. In: Proceedings of the 4th International Congress for Logic, Methodology and Philosophy of Science (Stud Logic Found Math 74), North Holland, Amsterdam, 1973 pp. 137-154
- ROBINSON A.: Standard and nonstandard number systems, Nieuw Arch Wisk, Ser 3 Vol. 21, 1973 pp. 115-133
- ROBINSON J.: Definability and decision problems in arithmetic, Journ. Symb. Log. Vol. 14, 1949 pp. 98-114
- ROBINSON J.: Existential definability in arithmetic, Trans Amer. Math. Soc. Vol. 72, 1952 pp. 437-449
- ROBINSON J.: The undecidability of exponential Diophantine equations. Notices of the Amer. Math. Soc. Vol. 7, 1960 pp. 75
- ROBINSON J.: The undecidability of exponential Diophantine equations. In: Logic, Methodology and Philosophy of Science, Proceedings of the 1960 International Congress, Stanford University Press, 1962 pp. 12-13 (Russian translation in *Matematičeskaja logika i ee primeneniya*, Mir 1965, 7-8)

- ROBINSON J.: Diophantine decision problems. In: *Studies in Number Theory*, MAA, Englewood Cliffs, 1969, pp. 76-116
- ROBINSON J.: Unsolvable Diophantine problems, *Proc. Amer. Math. Soc.* Vol. 22, 1969 pp. 534-538
- ROBINSON J.: Hilbert's tenth problem. In: *Institute on Number Theory*, Amer. Math. Soc.: Providence, 1971, pp. 191-194
- ROBINSON J.: Axioms for number theoretic functions. In: *Izbrannyje Voprosy Algebrы i Logiki: Sbornik Posvyasč. pamyati A. I. Mal'tseva*, Nauka: Novosibirsk, 1973 pp. 253-263
- ROBINSON J.: Solving Diophantine equations. In: *Logic, Methodology and Philosophy of Science IV*, North-Holland: Amsterdam, 1973, pp. 63-67
- ROBINSON R. M.: Arithmetical definitions in the ring of integers, *Proc. Amer. Math. Soc.* Vol. 2, 1951 pp. 279-284
- ROBINSON R. M.: An essentially undecidable axiom system. In: *Int Congr Math (II,6); 1950 Cambridge MA*, proceedings of the International Congress of Mathematicians, Amer. Math. Soc., 1952 pp. 729-730
- ROBINSON R. M.: Arithmetical representation of recursively enumerable sets, *Journ. Symb. Log.* Vol. 21, 1956 pp. 162-187
- ROBINSON R. M.: Restricted set-theoretical definitions in arithmetic. In: *Summaries of Talks Presented at the Summer Institute for Symbolic Logic*, Institute for Defense Analyses, Comm. Res. Division, 1957 pp. 139-140
- ROBINSON R. M.: Some representations of Diophantine sets, *Journ. Symb. Log.* Vol. 37, 1972 pp. 572-578
- ROGERS H. JR.: *Theory of recursive functions and effective computability*, McGraw-Hill, 1967, 482 pp.
- ROSE H. E.: *Subrecursive hierarchies*. PhD thesis, Princeton 1965.
- ROSE H. E.: *Subrecursion, Functions and Hierarchies*, Clarendon Press, Oxford, 1984
- ROSSER J. B.: Extensions of some theorems of Gödel and Church, *Journ. Symb. Log.* Vol. 1, 1936 pp. 87-91
- ROSSER J. B.: An informal exposition of proofs of Gödel and Church's theorem, *Journ. Symb. Log.* Vol. 4, 1939 pp. 53-60
- RYLL-NARDZEWSKI C.: The role of the axiom of induction in elementary arithmetic, *Fund. Math.* Vol. 39, 1952 pp. 239-263
- SAMBIN G.: Un'estensione del teorema di Loeb, *Rend Sem. Mat. Univ. Padova* Vol. 52, 1974 pp. 193-199
- SAMBIN G.: An effective fixed-point theorem in intuitionistic diagonalizable algebras (The algebraization of the algebras which express Theor. IX.), *Studia Logica* Vol. 35, 1976 pp. 345-361
- SAMBIN G.: Fixed points through the finite model property, *Studia Logica* Vol. 37, 1978 pp. 287-289
- SAMBIN G., VALENTINI S.: A modal calculus for a fragment of arithmetic, *Studia Logica* Vol. 39, 1980 pp. 245-256
- SAMBIN G., VALENTINI S.: The modal logic of provability. The sequential approach, *Journ. Philos. Log.* Vol. 11, 1982 pp. 311-342
- SAVAGE J. E.: *The Complexity of Computing*, Wiley, London, 1976, 391 pp.
- SAVITCH W. J.: Relationship between nondeterministic and deterministic tape complexities, *Journ. Comput. and System Sci.*, Vol. 4, 1970 pp. 177-192
- ŠAVRUKOV V. YU.: The logic of relative interpretability over Peano arithmetic (preprint), 1988
- SCANLON T. M.: The consistency of number theory via Herbrand's theorem, *Journ. Symb. Log.* Vol. 38, 1973 pp. 29-58
- SCARPELINI B.: Zwei unentscheidbare Probleme der Analysis, *Zeitschr. Math. Log. Grundl. Math.* Vol. 9, 1963 pp. 265-289

- SCHMERL J. H.: Peano models with many generic classes. *Pac. Journ. Math.* Vol. 46, 1973 pp. 523-536
- SCHMERL J. H.: Extending models of arithmetic, *Ann. Math. Logic* Vol. 14, 1978 pp. 89-109
- SCHMERL J. H.: Recursively saturated, rather classless models of Peano arithmetic. In: *Logic Year 1979-80. The University of Connecticut (Lect. Notes Math. 859)*, Springer-Verlag, 1981 pp. 268-282
- SCHMERL J. H.: Models of Peano arithmetic and a question of Sikorski on ordered fields, *Israel Journ. Math.* Vol. 50, 1985 pp. 145-159
- SCHMERL J. H.: Substructure lattices of models of Peano arithmetic. (abstract), *Journ. Symb. Log.* Vol. 51, 1986 p. 496
- SCHMERL J. H.: Substructure lattices of models of Peano arithmetic. In: *Logic Colloquium '84 (Studies in Logic and the Foundations of Mathematics 120)*, North-Holland, 1986 pp. 225-243
- SCHMERL J. H., SIMPSON S. G.: On the role of Ramsey quantifiers in first order arithmetic, *Journ. Symb. Log.* Vol. 47 1982 pp. 423-435
- SCHMERL U. R.: Eine von Reflexionsformeln erzeugte Feinstruktur über Erweiterungen der primitiv rekursiven Arithmetik. Dissertation, Heidelberg, 1978
- SCHMERL U. R.: A fine structure generated by reflection formulas over primitive recursive arithmetic. In: *Logic Colloquium '78 (Stud Logic Found Math 97)*, North Holland: Amsterdam, 1979 pp. 335-350
- SCHMERL U. R.: Iterated reflection principles and the ω -rule, *Journ. Symb. Log.* Vol. 47, 1982 pp. 721-733
- SCHMERL U. R.: Diophantine equations in a fragment of number theory. In: *Computation and proof theory (Lect. Notes Math. 1104)*, Springer-Verlag, 1984 pp. 389-398
- SCHMERL U. R.: Diophantine equations in a fragment of number theory. (abstract), *Journ. Symb. Log.* Vol. 50, 1985 pp. 275
- SCHMERL U. R.: Diophantine equations in fragments of arithmetic, *Annals Pure Appl. Logic* Vol. 38, 1988 pp. 135-170
- SCHMIDT D.: Associative ordinal functions, well partial orderings and a problem of Skolem, *Zeitschr. Math. Log. Grundle. Math.* Vol. 24, 1978 pp. 297-302
- SCHÜTTE K.: Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie, *Math. Ann.* Vol. 122, 1951 pp. 369-389
- SCHÜTTE K.: Beweistheorie. Springer-Verlag, 1960, 355 pp.
- SCHÜTTE K., SIMPSON S. G.: Ein in der reinen Zahlentheorie unbeweisbarer Satz über endliche Folgen von natürlichen Zahlen, *Arch. Math. Log.* Vol. 25, 1985 pp. 75-89
- SCHWARTZ D. G.: A free-variable theory of primitive recursive arithmetic, *Zeitschr. Math. Log. Grundle. Math.* Vol. 33 1987 pp. 147-157
- SCHWICHTENBERG H.: Eine Klassifikation der ε_0 -rekursiven Funktionen, *Arch. Math. Log.* vol. 13, 1970, pp. 204-236
- SCHWICHTENBERG H.: Proof theory: some applications of cut-elimination. In: *Handbook of Mathematical Logic*, Barwise ed. North-Holland: Amsterdam, 1977, pp. 867-896
- SCHWICHTENBERG H., WAINER S. S.: Infinite terms and recursion in higher types. In: *Proof Theory Symposion, Kiel 1974 (Lect. Notes Math. 500)*, Springer-Verlag, 1975 pp. 341-364
- SCOTT D. S.: On constructing models for arithmetic. In: *Infinitistic Methods. Proceedings of the Symposium on Foundations of Mathematics*, PWN: Warsaw, 1961 pp. 235-255
- SCOTT D. S.: Algebras of sets binumerable in complete extensions of arithmetic. In: *Recursive Function Theory*, Amer. Math. Soc.: Providence, 1962 pp. 117-121
- SEMENOV A. L.: On certain extensions of the arithmetic of addition of natural numbers, *Izv. Akad. Nauk SSSR, Ser. Mat.* Vol. 43, 1979 pp. 1175-1195, 9
- SEMENOV A. L.: On the definability of arithmetic in its fragments, *Doklady Akad. Nauk SSSR* Vol. 263, 1982 pp. 44-47

- SEMENOV A. L.: Logical theories of one-place functions on the set of natural numbers, *Izv. Akad. Nauk SSSR, Ser. Mat.* Vol. 47, 1983 pp. 623-658
- SEREMENT Z.: On automorphisms of resplendent models of arithmetic, *Zeitschr. Math. Log. Grundl. Math.* Vol. 30, 1984 pp. 349-352
- SHELAH S.: End extensions and numbers of countable models, *Journ. Symb. Log.* Vol. 43, 1978 pp. 550-562
- SHELAH S.: Models with second order properties II. trees with no undefined branches, *AML* Vol. 14, 1978 pp. 73-87
- SHELAH S.: On logical sentences in PA. In: *Proceedings of the Logic Colloquium '82*; Firenze, North Holland, Amsterdam, 1984 pp. 145-160
- SHEPHERDSON J. C.: Representability of recursively enumerable sets in formal theories, *Arch. Math. Log.* Vol. 5, 1960 pp. 119-127
- SHEPHERDSON J. C.: A non-standard model for a free variable fragment of number theory, *Bull. Acad. Polon. Sci.* Vol. 12, 1964 pp. 79-86
- SHEPHERDSON J. C.: Non-standard models for fragments of number theory. In: *The Theory of Models* (Stud. Logic Found. Math.), North Holland, Amsterdam, 1965 pp. 342-358
- SHEPHERDSON J. C.: The rule of induction in the free variable arithmetic based on $+$ and $*$, *Ann. Fac. Sci. Clermont* Vol. 35, 1967 pp. 25-31
- SHEPHERDSON J. C.: Weak and strong induction, *Amer. Math. Monthly* Vol. 76, 1969 pp. 989-1004
- SHOENFIELD J. R.: A relative consistency proof, *Journ. Symb. Log.* Vol. 19, 1954 pp. 21
- SHOENFIELD J. R.: Open sentences and the induction axiom, *Journ. Symb. Log.* Vol. 23, 1958 pp. 7-12
- SHOENFIELD J. R.: On a restricted ω -rule, *Bull. Acad. Polon. Sci.* Vol. 7, 1959 pp. 405-407
- SHOENFIELD J. R.: Degrees of models, *Journ. Symb. Log.* Vol. 25, 1960 pp. 233-237
- SHOENFIELD J. R.: Undecidable and creative theories, *Fund. Math.* Vol. 49, 1961 pp. 171-179
- SHOENFIELD J. R.: *Mathematical Logic*, Addison-Wesley, 1967, 344 pp.
- SHOSTAK R. E.: On the SUP-INF method for proving Presburger Formulas, *Journ. A. C. M.* Vol. 24, 1977 pp. 529-543
- SIEG W.: Fragments of arithmetic, *Annals Pure Appl. Logic* Vol. 28, 1985 pp. 33-71
- SIEG W.: A note on König's lemma. (abstract), *Journ. Symb. Log.* Vol. 50, 1985 pp. 276
- SIERPINSKI W.: 250 problems of elementary number theory, Warszawa, 1970
- SILVER J.: Harrington's version of the Paris's result. (manuscript)
- SIMMONS H.: Topological aspects of suitable theories, *Proc. Edinburgh Math. Soc.* Vol. 19, 1974 pp. 383-391
- SIMMONS H.: Each regular number structure is biregular, *Israel Journ. Math* Vol. 23, 1976 pp. 347-352
- SIMMONS H.: Existentially closed models of basic number theory. In: *Logic Colloquium '76*, North-Holland: Amsterdam 1977, pp. 325-369
- SIMMONS H.: The lattice of universal sentences modulo Peano arithmetic, *Rapp. Sem. Math. Pure, Univ. Catol. Louvain*, 1980 28 p.
- SIMPSON S. G.: Forcing and models of arithmetic, *Proc Amer. Math. Soc.* Vol. 43, 1974 pp. 193-194
- SIMPSON S. G.: Σ_1^1 and Π_1^1 transfinite induction. In: *Logic Colloquium '80*, North Holland, 1982 pp. 239-253
- SIMPSON S. G.: Set theoretic aspects of ATR_0 . In: *Logic Colloquium '80*, North Holland, 1982 pp. 255-271
- SIMPSON S. G.: Which set existence axioms are needed to prove Cauchy-Peano theorem for ordinary differential equations?, *Journ. Symb. Log.* Vol. 49, 1984 pp. 783-802
- SIMPSON S. G.: Nichtbeweisbarkeit von gewissen kombinatorischen Eigenschaften endlicher Bäume, *Arch. Math. Log.* Vol. 25, 1985 pp. 45-65

- SIMPSON S. G.: Unprovable theorems and fast-growing functions. In: Logic and Combinatorics, Proceedings of the Amer. Math. Soc. – SIAM Joint Summer Research Conference, Amer. Math. Soc.: Providence, 1987 pp. 359–394
- SIMPSON S. G.: Reverse mathematics, Dept. of Math. Research Rep., 1982, Penn State Univ.
- SIMPSON S. G.: Subsystems of second order arithmetic, book in preparation, Chapters II–IX distributed as Dept. of Math. Research Rep., 1986–90, Penn State Univ.
- SIMPSON S. G., SMITH R. L.: Factorization of polynomials and Σ_1^0 induction, *Annals Pure Appl. Logic* Vol. 31, 1986 pp. 289–306
- SKOLEM T.: Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen, *Skrifter utgit av videnskasselskapet i Kristiana* vol. I no. 4, 1920, pp. 1–36
- SKOLEM T. A.: Begründung der elementaren Arithmetik durch die rekurrerende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich, *Videnskaps Selskapet i Kristiana. Skrifter Utgit. 1 Matematisk-Naturvidenskabelig Klasse* Vol. 6, 1923 pp. 1–38
- SKOLEM T. A.: Über einige Satzfunktionen in der Arithmetik, *Skrifter Vitenskapssakademiet i Oslo*, I Vol. 7, 1930 pp. 1–28
- SKOLEM T. A.: Über die Unmöglichkeit einer Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems, *Norsk Matematisk Forenings Skrifter* Vol. 10, 1933 pp. 73–82
- SKOLEM T. A.: Über die Nichtcharakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vielen Aussagen mit ausschliesslich Zahlvariablen, *Fund. Math.* Vol. 23, 1934 pp. 150–161
- SKOLEM T. A.: Über die Zurückführbarkeit einiger durch Rekursionen definierter Relationen auf "arithmetische", *Acta Litt. scient. Szeged* Vol. 8, 1937 pp. 73–88
- SKOLEM T. A.: Peano's axioms and models of arithmetic. In: *Mathematical Interpretations of Formal Systems* (Stud Logic Found Math 10), North Holland, Amsterdam, 1955 pp. 1–14
- SKOLEM T. A.: A version of the proof of equivalence between complete induction and uniqueness of primitive recursion, *Kongelige Norske Videnskabers Selskabs. Forhandling.* Vol. 29, 1956 pp. 10–15
- SKOLEM T. A.: An ordered set of arithmetic functions representing the least epsilon-number, *Kongelige Norske Videnskabers Selskabs. Forhandling.* Vol. 29, 1956 pp. 54–59
- SKOLEM T.: *Selected works in logic* (ed. by J. E. Fenstad), Universitetsforlaget Oslo 1970, 732 pp.
- SLATER D.: Non-standard languages and their applications. Dissertation, Rutgers, 1972
- SMITH R., SIMPSON S. G.: Factorization of polynomials and Σ_1^0 induction. (abstract), *Journ. Symb. Log.* Vol. 51, 1986 pp. 497
- SMITH S. T.: Nonstandard syntax and semantics and full satisfaction classes for models of arithmetic. (abstract), *Journ. Symb. Log.* Vol. 51, 1986 pp. 497
- SMITH S. T.: Nonstandard characterizations of recursive saturation and resplendency, *Journ. Symb. Log.* Vol. 52 1987 pp. 842–863
- SMORYŃSKI C. A.: Applications of Kripke models. In: *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Springer-Verlag, 1973 pp. 324–391
- SMORYŃSKI C. A.: Investigations of intuitionistic formal systems by means of Kripke models. Dissertation, Chicago Circle, 1973
- SMORYŃSKI C. A.: The incompleteness theorems. In: *Handbook of Math Logic* (Stud Logic Found Math 90), North Holland: Amsterdam, 1977 pp. 821–865
- SMORYŃSKI C. A.: ω -consistency and reflection. In: *Colloque International de Logique* (Colloq Int CNRS), CNRS Inst. B. Pascal: Paris, 1977 pp. 167–181
- SMORYŃSKI C. A.: A note on the number of zeros of polynomials and exponential polynomials, *Journ. Symb. Log.* Vol. 42, 1977 pp. 99–106

- SMORYŃSKI C. A.: Beth's theorem and self-referential sentences. In: *Logic Colloquium 77* (Stud. Logic Found. Math. 96), North Holland: Amsterdam, 1978 pp. 253-261
- SMORYŃSKI C. A.: Avoiding self-referential statements. *Proceedings of the Amer. Math. Soc.* Vol. 70, 1978 pp. 181-184
- SMORYŃSKI C. A.: Calculating self-referential statements. I: Explicit calculations, *Studia Logica* Vol. 38, 1979 pp. 17-36
- SMORYŃSKI C. A.: Some rapidly growing functions, *Math. Intell.* Vol. 2, 1979 pp. 149-154
- SMORYŃSKI C. A.: Calculating self-referential statements, *Fund. Math.* Vol. 109, 1980 pp. 189-210
- SMORYŃSKI C. A.: Recursively saturated non-standard models of arithmetic, *Journ. Symb. Log.* Vol. 46, 1981 pp. 259-286
- SMORYŃSKI C. A.: Calculating self-referential sentences: Guaspari sentences of the first kind, *Journ. Symb. Log.* Vol. 46, 1981 pp. 329-344
- SMORYŃSKI C. A.: Fifty years of self-reference, *Notre Dame Journ. Formal Log.* Vol. 22, 1981 pp. 357-374
- SMORYŃSKI C. A.: Cofinal extensions of nonstandard models of arithmetic, *Notre Dame Journ. Formal Log.* Vol. 22, 1981 pp. 133-144
- SMORYŃSKI C. A.: Elementary extensions of recursively saturated models of arithmetic, *Notre Dame Journ. Formal Log.* Vol. 22, 1981 pp. 193-203
- SMORYŃSKI C. A.: Fixed point algebras, *Bull. Amer. Math. Soc.* Vol. 6, 1982 pp. 317-356
- SMORYŃSKI C. A.: Nonstandard models and constructivity. In: *The L. E. J. Brouwer Centenary Symposium* (Noordwijkerhout, 1981), North-Holland, Amsterdam-New York, 1982 pp. 459-464
- SMORYŃSKI C. A.: A note on initial segment constructions in recursively saturated models of arithmetic, *Notre Dame Journ. Formal Log.* Vol. 23, 1982 pp. 393-408
- SMORYŃSKI C. A.: The varieties of arboreal experience. *Math. Intell.* Vol. 4, 1982 pp. 182-189
- SMORYŃSKI C. A.: The finite inseparability of the first-order theory of diagonalisable algebras, *Studia Logica* Vol. 41, 1982 pp. 347-349
- SMORYŃSKI C. A.: Commutativity and self reference, *Notre Dame Journ. Formal Log.* Vol. 23, 1982 pp. 443-452
- SMORYŃSKI C. A.: Back-and-forth inside a recursively saturated model of arithmetic. In: *Logic Colloquium '80* (Stud. Logic Found. Math. 108), North Holland: Amsterdam, 1982 pp. 273-278
- SMORYŃSKI C. A.: Recursively saturated models of arithmetic, *Addendum, Journ. Symb. Log.* Vol. 47, 1982 pp. 493
- SMORYŃSKI C. A.: Modal logic and self-reference. In: *Handbook of Philosophical Logic* (vol. 2), Reidel: Dordrecht, 1984 pp. 441-496
- SMORYŃSKI C. A.: Lectures on nonstandard models of arithmetic. In: *Proceedings of the Logic Colloquium '82* (Stud. Logic Found. Math. 112), North Holland: Amsterdam, 1984 pp. 1-70
- SMORYŃSKI C. A.: Self-reference and Modal logic, Springer-Verlag, 1985, 333 pp.
- SMORYŃSKI C. A.: Nonstandard models and related developments. In: *Harvey Friedman's Research on the Foundations of Mathematics*, North Holland: Amsterdam, 1985 pp. 179-229
- SMORYŃSKI C. A.: Quantified modal logic and self-reference, *Notre Dame Journ. Formal Log.* Vol. 28, 1987 pp. 356-370
- SMORYŃSKI C. A.: Logical number theory I - an introduction, Springer-Verlag 1991
- SMORYŃSKI C. A.: An asymptotic formula for a logico-combinatorial problem, preprint
- SMORYŃSKI C. A., STAVI J.: Cofinal extension preserves recursive saturation. In: *Model Theory of Algebra and Arithmetic* (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 338-345
- SMULLYAN R. M.: Languages in which self reference is possible, *Journ. Symb. Log.* Vol. 22, 1957 pp. 55-67

- SMULLYAN R. M.: Theory of Formal Systems, *Annals of Math. Stud. No. 47*, Princeton Univ. Press, 1961, 142 pp.
- SMULLYAN R. M.: Modality and self-reference. In: *Intensional Mathematics (Stud. Logic Found. Math. 113)*, North Holland: Amsterdam, 1985 pp. 191–211
- SMULLYAN R. M.: Some principles related to Loeb's theorem. In: *Intensional Mathematics (Stud. Logic Found. Math. 113)*, North Holland: Amsterdam, 1985 pp. 213–230
- SMULLYAN R. M.: Fixed points and self-reference, *Int Journ. Math. & Math. Sci. Vol. 7*, 1984 pp. 283–289
- SOARE R. I.: *Recursively enumerable sets and degrees*. Springer-Verlag 1987.
- SOLOMON M. K.: The Gödel speed-up phenomenon. Dissertation, Stevens Inst. of Tech, 1975
- SOLOMON M. K.: Some results on measure independent Gödel speed-ups, *Journ. Symb. Log. Vol. 43*, 1978 pp. 667–672
- SOLOVAY R. M.: Provability interpretations of modal logic, *Israel Journ. Math. Vol. 25*, 1976 pp. 287–304
- SOLOVAY R. M.: Explicit Henkin sentences, *Journ. Symb. Log. Vol. 50*, 1985 pp. 91–93
- SOLOVAY R. M.: Infinite fixed-point algebras. In: *Recursion Theory. Proceedings of the A. M. S. – A. S. L. Summer Institute*, Amer. Math. Soc: Providence, 1985 pp. 473–486
- SOLOVAY R. M.: Injecting inconsistencies into models of *PA*, *Annals of Pure and Appl. Log. Vol. 44*, 1989, pp. 101–132
- SOMMER R.: Ordinal arithmetic in Δ_0 . In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 320–363
- SOPRUNOV S. F.: Strong nonstandard models of arithmetic, *Doklady Akad. Nauk SSSR Vol. 223*, 1975 pp. 576–577
- SOPRUNOV S. F.: Initial segments of nonstandard arithmetics, *Doklady Akad. Nauk SSSR Vol. 223*, 1975 pp. 576–577
- SOPRUNOV S. F.: Denumerable non-standard models of arithmetic. In: *Issledovaniya po Teorii Množestv i Neklassičeskim Logikam. Sbornik Trudov*, Nauka: Moskva, 1976 pp. 157–173
- SOPRUNOV S. F.: Lattices of non-standard arithmetics. In: *Issledovaniya po Neklassičeskim Logikam i Teorii Množestv*, Nauka: Moskva, 1979 pp. 146–173
- SORNIKOV YA. A.: On quantifier-free ε_0 -recursive arithmetic, *Vestnik Mosk. Univ. Matematika Vol. 32*, 1977 pp. 19–25
- SPECKER E.: Eine Verschärfung des Unvollständigkeitssatzes der Zahlentheorie, *Bull. Acad. Polon. Sci. Vol. 5*, 1957 pp. 1041–1045
- SPENCER J.: Large numbers and unprovable theorems, *Amer. Math. Monthly Vol. 90*, 1983 pp. 669–675
- STEIN M.: Interpretations of Heyting's arithmetic – an analysis by means of a language with set symbols, *Ann. Math. Logic Vol. 19*, 1980 pp. 1–31
- STEPANOV V. I.: Self-referential propositions for standard and Rosser provability predicates, *Usp. Mat. Nauk Vol. 36* 1981 pp. 227–228
- STEPANOV V. I.: Second-order arithmetic and the consistency of first-order theories, *Usp. Mat. Nauk Vol. 37*, 1982 pp. 179–180
- STIHI T.: Forcing and Peano arithmetic. Dissertation, 1976
- STIHI T.: Induction and complete sequences of conditions in the sense of Cohen, *Rev. Roum. Math. Pures et Appl. Vol. 23*, 1978 pp. 129–133
- STOCKMEYER L.: The polynomial time hierarchy, *Theor. Computer Science Vol. 3*, 1976 pp. 1–22
- SUNDHOLM B. G.: The ω -rule; a survey, Bachelor's thesis, Oxford, 1978
- ŠVEJDAR V.: Degrees of interpretability, *Comm. Math. Univ. Carol. Vol. 19*, 1978 pp. 789–813
- ŠVEJDAR V.: A sentence that is difficult to interpret, *Comm. Math. Univ. Carol. Vol. 22*, 1981 pp. 661–666

- ŠVEJDAR V.: Modal analysis of generalized Rosser sentences, *Journ. Symb. Log.* Vol. 48, 1983 pp. 986–999
- SZELEPCZÉNYI R.: The method of forcing for nondeterministic automata, *Bull. European Assoc. for Theor. Comp. Sci.*, Vol. 33, 1987 pp. 96–100
- SZMIELEW W., TARSKI A.: Mutual interpretability of some essentially undecidable theories. In: *Int Congr Math. (II,6)*; 1950 Cambridge MA, USA [1952] *Proceedings of the International Congress of Mathematicians*, Amer. Math. Soc. 1952 pp. 734
- TAIT W. W.: The substitution method, *Journ. Symb. Log.* Vol. 30, 1965 pp. 175–192
- TAIT W. W.: Functionals defined by transfinite recursion, *Journ. Symb. Log.* Vol. 30, 1965 pp. 155–174
- TAKAHASHI M.: A foundation of finite mathematics, *Publ. Res. Inst. Math. Sci. Kyoto Univ.* Vol. 12, 1977 pp. 577–708
- TAKAHASHI M.: An abstract form of Gödel's theorem on consistency and Löb's, *Sci. Reports Tokyo Kyoiku Daigaku, Sec. A* Vol. 13, 1977 pp. 81–85
- TAKEUTI G.: Proof theory, North-Holland P. C, 1975, 372 pp. (second edition 1980)
- TAKEUTI G.: Consistency proofs and ordinals. In: *Proof Theory Symposium*, Kiel 1974 (Lect. Notes Math. 500), Springer-Verlag, 1975 pp. 365–369
- TAKEUTI G.: Work of Paul Bernays and Kurt Gödel. In: *Logic, Methodology and Phil. of Science VI*, North-Holland P. C. 1982 pp. 77–85
- TAKEUTI G.: Some relations among systems for bounded arithmetic, *Ann. Pure Appl. Logic* Vol. 39, 1988 pp. 75–104
- TAKEUTI G.: Bounded arithmetic and truth definition, *Annals Pure Appl. Log.* Vol. 39, 1988 pp. 75–104
- TAKEUTI G.: S_3^1 and $V_2^1(BD)$, *Arch. Math. Log.* Vol. 29, 1990 pp. 149–169
- TAKEUTI G.: Some relations among systems of bounded arithmetic. In: *Heyting: Mathematical Logic*, P. Petkov ed., Plenum Press, 1990, pp. 139–154
- TAKEUTI G.: Sharply bounded arithmetic and the function $\alpha \cdot 1$. In: *Workshop in Logic and Computation*, Amer. Math. Soc. Contemporary Math. 106, 1990, pp. 281–288
- TAKEUTI G.: A second order version of S_2 and U_2^1 , *Journal of Symb. Logic* 56, 1991, pp. 1038–1063
- TAKEUTI G.: RSUV Isomorphism. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajíček, Oxford Univ. Press 1992, pp. 364–386
- TARSKI A.: Pojęcie prawdy w językach nauk dedukcyjnych, *Prace Tow. Naukowego Warszawskiego* vol. III, no. 34, 1933, 116 pp.
- TARSKI A.: Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Phil.* Vol. 1, 1936 pp. 261–405
- TARSKI A.: Logic, semantics, metamathematics: Papers from 1923 to 1938 (translated by J. H. Woodger). Clarendon Press, Oxford 1956, 471 pp.
- TARSKI A. MOSTOWSKI A. ROBINSON R. M.: Undecidable theories, North Holland: Amsterdam 1953, 98 pp.
- TENNENBAUM S.: Non-archimedean models for arithmetic. In: *Notices Amer. Math. Soc.*, 1959 pp. 270
- TODA S.: On the computational power of PP and $\oplus$ PP, 30-th Symp. on Foundations of Comp. Sci., 1989 pp. 514–519
- TRAHTENBROT B. A.: The impossibility of an algorithm for decision problem for finite domains, *Doklady Akad. Nauk SSSR* Vol. 70, 1950 pp. 569–572
- TRAHTENBROT B. A.: On recursive separability, *Doklady Akad. Nauk SSSR* Vol. 88, 1953 pp. 953–956
- TSUBOI A.: On M -recursively saturated models of arithmetic, *Tsukuba Journ. Math.* Vol. 6, 1982 pp. 305–318
- TURING A. M.: Systems of logic based on ordinals, *Proc. London Math. Soc. (2)* Vol. 45, 1939 pp. 161–228
- TVERSKOY A. A.: A sequence of combinatorial judgments which are independent of Peano Arithmetic, *Moscow University Math. Bull.* Vol. 35, 1980 pp. 6–13

- TVERSKOY A. A.: Investigation of recursiveness and arithmeticity of signature functions in nonstandard models of arithmetic, *Dokl. Acad. Nauk USSR* Vol. 262, 1982, pp. 1325-1328
- VON DER TWER T.: Some remarks on the mathematical incompleteness of Peano's arithmetic found by Paris and Harrington. In: *Set theory and model theory* (Bonn, 1979) (Lect. Notes Math. 872), Springer-Verlag, 1981 pp. 157-174
- VON DER TWER T.: On the strength of several versions of Dirichlet's ("pigeon-hole") principle in the sense of first-order logic, *Arch. Math. Log.* Vol. 21, 1981 pp. 69-76
- URSINI A.: A sequence of theories for arithmetic whose union is complete, *Rend. Semin. Mat. Univ. Padova* Vol. 57, 1977 pp. 75-92
- USPENSKIY V. A.: The Gödel theorem and the theory of algorithms, *Usp Mat Nauk* Vol. 8/4(56), 1953 pp. 176-178
- USPENSKIY V. A.: Theorem of Gödel and theory of algorithms, *Doklady Akad. Nauk SSSR* Vol. 91, 1953 pp. 737-740
- VAANANEN J.: A new incompleteness in arithmetic, *Arkhimedes* Vol. 31, 1979 pp. 30-37
- VALENTINI S. SOLITRO U.: The modal logic of consistency assertions of Peano arithmetic, *Zeitschr. Math. Log. Grundle. Math.* Vol. 29, 1983 pp. 25-32
- VARDANYAN V. A.: Arifmetičeskaya složnost predikatnykh logik dokazuemosti i ich fragmentov, *Doklady Akademii Nauk SSSR* Vol. 288, 1986 pp. 11-14
- VAUGHT R. L.: Sentences true in all constructive models, *Journ. Symb. Log.* Vol. 25, 1960 pp. 39-53
- VAUGHT R. L.: On a theorem of Cobham concerning undecidable theories. In: *Logic, Methodology and Philosophy of Science*, Stanford, 1962, pp. 14-25
- VAUGHT R.: On axiomatizability by a schema, *Journ. Symb. Log.* vol. 32, 1967, pp. 473-479
- VERBRUGGE L. C.: Does Solovay's completeness theorem extend to bounded arithmetic? (Master's thesis), Amsterdam, 1988
- VERBRUGGE R.: Feasible Interpretability. In: *Arithmetic, Proof Theory and Computational Complexity*, eds. P. Clote and J. Krajčček, Oxford Univ. Press 1992, pp. 389-400
- VISSEER A.: Numerations, lambda-calculus & arithmetic. In: *To H. B. Curry: essays on combinatory logic, lambda calculus and formalism*, Academic Press, London, 1980 pp. 259-284
- VISSEER A.: On the completeness principle: a study of provability in Heyting's arithmetic and extensions, *Ann. Math. Logic* Vol. 22, 1982 pp. 263-295
- VISSEER A.: The provability logics of recursively enumerable theories extending Peano arithmetic at arbitrary theories extending Peano arithmetic, *Journ. Philos. Logic*, Vol. 13, 1984 pp. 97-113
- VISSEER A.: Evaluation, provably deductive equivalence in Heyting's arithmetic of substitution instances of propositional formulas, *Logic group preprint series No 4 Univ. Utrecht*, 1985, 58 p.
- VISSEER A.: Preliminary notes on interpretability logic, *Logic group preprint series Univ. Utrecht*, 1988, 37 p.
- VISSEER A.: Peano's smart children: A provability logical study of systems with builtin consistency, *Notre Dame Journ. Formal Logic* 30, 1989, pp. 161-196
- VISSEER A.: Interpretability logic. In: *Mathematicae Logic*, Petkov ed., Plenum Press, 1990, pp. 175-208
- VISSEER A.: An inside view of EXP, *Journ. of Symbolic Logic* 57, 1992, pp. 131-165
- VOPĚNKA P.: A new proof of the Gödel's result on non-provability of consistency, *Bull. Acad. Polon. Sci.* Vol. 14, 1966 pp. 111-116
- WAGNER K., WECHSUNG G.: *Computational Complexity*, VEB Deutscher Verlag der Wissenschaft, Berlin, 1986, 551 pp.
- WAINER S.: A classification of ordinal recursive functions, *Arch. Math. Log.* vol. 13, 1970, pp. 136-153
- WAINER S.: Ordinal recursion and a refinement of the extended Grzegorzczuk hierarchy, *Journ. Symb. Log.* vol. 37, 1972, 281-292

- WANG H.: The non-finitizability of impredicative principles, *Proc. Nat. Acad. Sci. USA* Vol. 36, 1950 pp. 479-484
- WANG H.: Remarks on the comparison of axiom systems, *Proc. Nat. Acad. Sci. USA* Vol. 36, 1950 pp. 448-453
- WANG H.: Arithmetic translation of axiom systems, *Trans Amer. Math. Soc* Vol. 71, 1951 pp. 283-293
- WANG H.: Arithmetic models for formal systems, *Methodos* Vol. 3, 1951 pp. 217-232
- WANG H.: Truth definitions and consistency proofs, *Trans Amer. Math. Soc* Vol. 73, 1952 pp. 243-275
- WANG H.: Undecidable sentences generated by semantic paradoxes, *Journ. Symb. Log.* Vol. 20, 1955 pp. 31-43
- WANG H.: The axiomatization of arithmetic, *Journ. Symb. Log.* Vol. 22, 1957 pp. 145-158
- WANG H.: The arithmetization of metamathematics. In: Wang H.: *Survey of Mathematical Logic*, North Holland: Amsterdam, 1963 pp. 334-361
- WANG P. S.: The undecidability of the existence of zeros of real elementary functions, *Journ. A. C. M.* Vol. 21, 1974 pp. 586-589
- WEGENER I.: *The Complexity of Boolean Functions*, Wiley/Teubner, New York and Stuttgart, 1987, 457 pp.
- WILKIE A. J.: On models of arithmetic - answer to two problems raised by H. Gaifman, *Journ. Symb. Log.* Vol. 40, 1975 pp. 41-47
- WILKIE A. J.: On the theories of end extensions of models of arithmetic. In: *Set theory and Hierarchy theory V. Proceedings of the 3rd Conference on Set Theory and Hierarchy theory* (Lect. Notes Math. 619), Springer-Verlag, 1977 pp. 305-310
- WILKIE A. J.: On models of arithmetic having non-modular substructure lattices, *Fund. Math.* Vol. 95, 1977 pp. 223-237
- WILKIE A. J.: Some results and problems on weak systems of arithmetic. In: *Logic colloquium '77*, Proc. Wroclaw 1977, *Stud. Logic Found. Math.* Vol. 96, 1978 pp. 285-296
- WILKIE A. J.: Applications of complexity theory to Σ_0 -definability problems in arithmetic. In: *Model Theory of Algebra and Arithmetic* (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 363-369
- WILKIE A. J.: On discretely ordered rings in which every definable ideal is principal. In: *Model Theory and Arithmetic Comptes Rendus d'une Action Thématique Programmée du C. N. R. S. sur la Théorie des Modèles et l'Arithmétique*, Paris, France, 1979/80 (Lect. Notes Math. 890), Springer-Verlag, 1981 pp. 297-303
- WILKIE A. J.: On core structures for Peano arithmetic. In: *Logic colloquium '80*, *Stud. Logic Found. Math.* 108, 1982 pp. 311-314
- WILKIE A. J.: On discretely ordered rings in which every definable ideal is principal. In: *Model theory and arithmetic* (Lect. Notes Math. 890), Springer-Verlag, 1985 pp. 197-303
- WILKIE A. J.: A model-theoretic proof of Buss's characterization of the polynomial time computable functions, manuscript, 1985
- WILKIE A. J.: On sentences interpretable in systems of arithmetic. In: *Logic Colloquium '84* (Studies in Logic and the Foundations of Mathematics 120), North-Holland, 1986 pp. 329-342
- WILKIE A. J.: On schemes axiomatizing arithmetic. In: *Proceedings ICM*, Berkeley, 1986 pp. 331-337
- WILKIE A. J., PARIS J. B.: Some results on bounded induction. In: *Proceedings of the second Easter conference on model theory* (Wittenberg, 1984), Humboldt Univ. Berlin,, 1984 pp. 223-228
- WILKIE A. J., PARIS J. B.: On the scheme of induction for bounded arithmetic formulas, *Annals Pure Appl. Logic* Vol. 35, 1987 pp. 261-302
- WILMERS G.: Minimally saturated models. In: *Model Theory of Algebra and Arithmetic* (Lect. Notes Math. 834), Springer-Verlag, 1980 pp. 370-380

- WILMERS G.: Bounded existential induction, *Journ. Symb. Log.* Vol. 50, 1985 pp. 72–90
- WOODS A.: Some problems in logic and number theory and their connections. Dissertation, Manchester, 1981
- WOODS A.: Bounded arithmetic formulas and Turing machines of constant alternation. In: *Logic Colloquium '84* (Studies in Logic and the Foundations of Mathematics 120), North-Holland, 1986 pp. 355–377
- WRATHALL C.: Rudimentary predicates and relative computation, *SIAM Journ. Comput.*, Vol. 7, 1978 pp. 194–209
- YASUGI M.: Gentzen reduction revisited, *Publ. RIMS Kyoto Univ.* Vol. 16, 1980 pp. 1–33
- YOUNG P.: Gödel theorems, exponential difficulty and undecidability of arithmetic theories: an exposition. In: *Recursion Theory. Proceedings of the A. M. S. – A. S. L. Summer Institute* (Proc. Symp. Pure Math. 42), Amer. Math. Soc: Providence, 1985 pp. 503–522
- YUKAMI T.: A theorem on the formalized arithmetic with function symbols $'$ and $+$, *Tsukuba Journ. Math.* Vol. 1, 1977 pp. 195–211
- YUKAMI T.: A note on formalized arithmetic with function symbols $'$ and $+$, *Tsukuba Journ. Math.* Vol. 2, 1978 pp. 69–73
- ZAHN G.: On the additive structure of non-standard models of arithmetic. Dissertation, Univ. S. Carolina, 1971
- ŽÁK S.: A Turing machine hierarchy, *Theor. Computer Science* Vol. 26, 1983 pp. 327–333
- ZBIERSKI P.: Indicators and incompleteness of Peano arithmetic, *Acta Cient. Venezolana* Vol. 31, 1980 pp. 487–495
- ZBIERSKI P.: Nonstandard interpretations of higher order theories, *Fund. Math.* Vol. 112, 1981 pp. 175–186

Index of Terms

- Absolute equality 149
- Accepted – in space 277
- Accepted – in time 277
- Acceptor 277
- Arithmetical hierarchy 13
- Arithmetization of metamathematics 20
- Axiomatic theory 7
- Axioms, special 7
- Axioms, logical 7

- Bertrand's postulate 50
- Binary search 339
- Bounded arithmetic 271
- Bounded collection axiom 328
- Bounded prenex normal form 76
- Bounded recursion 335
- Buss's witnessing theorem 349

- Cardinality 42
- Closure 337
- Coding finite sets and sequences 37
- Coding of finite sequences 17
- Collection 44
- Collection axiom given 61
- Comprehension 45
- Computation of a Turing machine 330
- Concatenation 19
- Connectives 6
- Conservative extension 148
- Consistency of a theory 376
- Consistency statement 163
- Consistency, bounded 411
- Consistency, cut-free 376
- Consistency, Herbrand 376
- Consistency, restricted 376
- Consistent 8
- Consistent 99
- Constants 6
- Contain a theory 150

- Counterexample 355
- Counting problem 409
- Craig's theorem 166
- Cut 215, 218, 367, 407
- Cut (definable) 172
- Cut, extendable 237
- Cut, proper 218
- Cut, Ramsey 237
- Cut, semiregular 235
- Cut, short 234
- Cut, unbounded 236
- Cut-rank 373

- Decidable 23
- Decidable 99
- Deduction rules 7
- Deductively equivalent 166
- Definable 7
- Definable element 222
- Definable subset of a model 215
- Definable ultrapower 216
- Defines a total function 47
- Designated variables 125
- Diagonal lemma 158
- Divide 33
- Dot notation 14
- Doubly conservative 197

- Elementary substructure 219
- End-extension 215
- Envelope 246
- Equality 6
- Equality absolute 6
- Essentially reflexive 168
- Evaluation 6
- Evaluation of variables 103
- Even 34
- Exponential time 278
- Exponentiation 18, 36

- Feasible solution 353
- Finite axiom of choice 67
- Finite axiomatizability 350
- Finite Ramsey theorem 112
- Fixed-point theorem 158
- Flexible formula 162
- Formula 6
- Formula, atomic 6
- Formula, cut formula 373
- Formula, inductive 367
- Formula, middle 377
- Formula, provable 7
- Formula, rudimentary 312
- Formula, side formula 373
- Formula, true 7
- Free algebra 21
- Function symbols 6

- Gödel sentence 160
- Gödel's second incompleteness theorem 164
- Gödel-Rosser incompleteness theorem 161
- Grammar 315
- Grammar, context-free 315
- Greatest common divisor 35

- H-Rosser formula 190
- Height of an ordinal 133
- Herbrand proof 179
- Herbrand variant 9
- Herbrand's theorem 10
- Herbrand's theorem 101
- Hereditarily conservative 197
- Hereditarily minimal prehomogeneous 115
- Hilbert-Ackermann's theorem 101
- Homogeneous 112

- Indicator 247
- Induction principle 5
- Induction, parameter-free 407
- Infinite Ramsey theorem 112
- Instance 62
- Interpretation 148
- Interpretation, faithful 150, 405
- Interpretation, global 367
- Interpretation, local 367
- Interpretation, parametrical 149
- "it's snowing"-it's snowing 59

- Language 6, 98, 315
- Language of arithmetic 12
- Large set 139
- Lattice of interpretability types 404

- Least number axiom 61
- Least number principle 5, 35
- Length 325
- Length of proof 373
- Length optimal solution 355
- Limit theorem 90
- Lindström's fixed point theorem 198
- Lindström's second fixed point theorem 208
- Linear space 278
- Linear time 278
- Linear time hierarchy 283
- Loeb's theorem 165
- Logarithmic space 278
- Logical axioms 7
- Los's lemma 217
- Low arithmetized completeness theorem 104
- Low basis theorem 92

- Matiyasevič's theorem 97
- Middle sequent theorem 377
- Minimal prehomogeneous 115
- Modal logic 405
- Model 6, 8, 102
- Model, standard 12
- Multi-exponentially bounded 175

- Narrowly branching 115
- Nepomnjaščij's theorem 285
- Nondeterministic polynomial time 279
- NP-complete 280
- Number, even 34
- Number, odd 34
- Number, prime 33
- Numeral, dyadic 318

- Odd 34
- Operation, suitable 249
- Optimization problem 353
- Order induction axiom 61
- Order-type of the universe 81
- Ordinal isolated 134
- Ordinal limit 134
- Overspill 219

- Pairing 34
- Parikh's theorem 272
- Paris sequence 126, 249
- Paris sequences of the first kind 250
- Paris sequences of the second kind 253
- Paris-Harrington
- Principle 112
- Partial satisfaction 56, 79

- Peano arithmetic 28
- Piecewise coded 63, 78
- Pigeon-hole principle 42, 67
- Polynomial (time) hierarchy 283
- Polynomial advice 354
- Polynomial advice 356
- Polynomial space 278
- Polynomial time 278
- Polynomially reducible 280
- Pre-ordinal 133
- Predicates 6
- Prehomogeneous 115
- Prehomogeneous 116
- Prenex normal form 8
- Prime 33
- Prime decomposition 50
- Primitive recursive functions 10
- Principle of alfa-large intervals 133
- Proof 372
- Proof line 372
- Proof, regular 377
- Provability conditions 163
- Provability in theory 375
- Provability, cut-free 375
- Provability, Herbrand 375
- Provability, restricted 375
- Provable 7
- Provably recursive 246
- Provably recursive 48
- Provably total 245
- Provably total 48
- Pure extension 168

- Quantifier 6
- Quantifier, sharply bounded 322

- Recursive functions 10
- Recursively saturated model 407
- Reflection principle 404
- Reflexive 168
- Regular tree 133
- Regularity 67
- Relatively large 112
- Replacement axiom 328
- Robinson arithmetic 28
- Rosser sentence 161

- Satisfaction relation 7
- Schwichtenberg-Wainer hierarchy 132, 140
- Second order arithmetic 151
- Second order systems 410
- Self-prover 211
- Self-reference 158

- Self-referential pairs 160
- Sequence 325
- Sequential 151
- Set, definable 7
- Sheperdson-Smorynski's fixed point theorem 198
- Skolem axiom 8
- Skolem function 337
- Sound 27
- Space 277
- Space constructible 288
- Special axiom 181
- Special axioms 7
- Speed up 211
- Standard model 12, 214, 234
- Standard system 407
- Strong collection 67
- Strong indiscernibles 125
- Strong replacement 329
- Strong Skolem property 231
- Subcut 367
- Substitution 314
- Subtheory 148
- Successor induction axiom 61

- Tarski's conditions for satisfaction 6
- Terms 6
- Theorem, Gödel incompleteness 386
- Theorem, Herbrand's 10
- Theorem, Matiyasevič(-Robinson-Davis-Putman) 19
- Theory 99
- Theory, decidable 23
- Theory, sound 27, 155
- Time 277
- Time constructible 288
- Time hierarchy 292
- Transducer 277
- Transfinite progression of theories 404
- True 7
- Truth definition 361
- Turing machine 277
- Turing machine – nondeterministic 279
- Turing machine – oracle 282

- Unbounded 81
- Underspill 219

- Value of a term 6
- Variables 6

- Witness 320
- Witness comparison 155

Index of Symbols

- $Subst(\varphi, x, t)$ 6
 $M \models \varphi[e]$ 6
 $sk(\Phi)$ 8
 $He(\Phi)$ 9
 Add 11
 $Mult$ 11
 Exp 11, 272
 $Th(N)$ 13
 Σ_n 13
 Π_n 13
 Δ_n 13
 OP 14
 Seq 17, 43, 296
 lh 17, 43, 310, 325
 $z = (x)_y$ 17, 43, 309, 325
 $s \smallfrown t$ 19, 43, 310
 Q 28
 PA 28
 $I_{open}, I\Sigma_0, I\Sigma_1$ 29, 268, 269
 $I\Sigma_0(exp)$ 37
 bit 38, 305, 324
 Sat_0 56
 $Sat_{\Sigma, n}$ 58
 $Sat_{\Pi, n}$ 58
 $I\Sigma_n$ 62
 $L\Sigma_n$ 62
 $B\Sigma_n$ 62
 $B\Pi_n$ 62
 $I\Delta_n$ 62
 LL_n 85
 $Term^\bullet$ 99, 312, 313
 $Proof_T^\bullet, Pr_T^\bullet, Con_T^\bullet$ 99, 163, 373, 375, 376
 $[X]^u$ 111
 $F : [X]^u \rightarrow a$ 112
 $\omega \rightarrow (\omega)_a^u$ 112
 $[x, y] \rightarrow_\star (q)_z^u$ 112
 $PH(u, z)$ 112
 $\omega \rightarrow (\omega)_z^u(\Gamma_1, \Gamma_2)$ 113
 $Arrow(n, i, j)$ 119
 $Arrow^0(n, i, j)$ 119
 $PH(u, z)$ 121
 $(PH)_u$ 121
 (PH) 121
 $\varphi \upharpoonright$ 125
 $\mathcal{O}_{x, y}$ 134
 $\preceq_{x, y}$ 134
 $\alpha \rightarrow_x \beta$ 135
 $L(\omega_k^n, \Sigma_m)$ 138
 $[x, y]$ 141
 $(W)_u$ 143
 $SEQ(z, u)$ 151
 ACA_0 151
 $|x|$ 172, 321
 2_y^x 172
 $\omega_y(x)$ 172
 $HProof_T^\bullet, HPr_T^\bullet, HCon_T^\bullet$ 179, 376
 Q_k -formulas 180
 $\Delta \prec^\# \nabla$ 196
 $Intp_T$ 206
 $Interp_T^\bullet$ 206
 $M \models \varphi$ 215
 $M \prec_k K$ 219
 $M \subseteq_c K$ 219
 $M \prec_{e, k} K$ 219
 $Rams(n, m)$ 236
 Σ_1^G 242
 P^- 268
 Σ_0^f 271

- $L_0(f)$ 271
 Ω_i 272
Superexp 272
 $I\Sigma_0 + \Omega_1$ 272
 $I\Sigma_0 + Exp$ 272
 $I\Sigma_0 + Superexp$ 272
 E_n 274
 IE_i 275
 $\Sigma_0\text{-PHP}$ 276
 $Time(f(n))$ 278
 $Space(f(n))$ 278
LinTime 278
 P 278
 E 278
 Exp 278
LogSpace 278
LinSpace 278
PSpace 278
 E^2_* 279
 NP 279
 co 280
CLIQUE 281
HAMILTONIAN GRAPHS 281
 Σ^{lin}_i 283
LinH 283
 Δ^p_i 283
 Σ^p_i 283
 PH 283
 $\square^p_i$ 283
 $\exists^{lin}_i$ 283
 $\forall^{lin}_i$ 283
 $\exists^p$ 283
 $\forall^p$ 283
 Σ^N_0 284
nuon 284, 304, 306
 $TimeSpace(f(n), g(n))$ 286
 Σ^f_i 292
 $TH(f(n))$ 292
 Pow 295, 324
 $lpw(x)$ 296
 $x \in (u, v)$ 296
 $Exp(x, y, z)$ 299
 $Exseq(x, s)$ 299
 $Exp(x, y, z)$ 301
 $\omega(x)$ 303
 $Exp_1(x, y, z)$ 303
 sgm 305
Nuonseq 305
nuon₁ 306
 $lf(x)$ 306
 $u \Rightarrow w$ 316
 $\Rightarrow$ 316
 $x \# y$ 321
 L_2 321
BASIC 321
 Σ^b_i 322
 Π^b_i 322
 Δ^b_i 323
 $\Sigma^b_0(\Sigma^b_i)$ 323
PIND 323
LIND 323
LMIN 323
 S^i_2 323
 T^i_2 323
 S_2 323
 T_2 323
 $\max$ 325
 bound 325
 msp 325
 $I\Delta^b_i$ 326
BB 328
 $\hat{L}_2$ 329
 $\hat{\Sigma}^b_i$ 329
 $\hat{\Pi}^b_i$ 329
 $\text{strict}\Sigma^b_i$ 329
Comp 330
 μ_i 336
 ν_i 336
 $\exists\Pi^b_i$ 343
 $\square^p_i/poly$ 356
Value 363
Var 363
val 364
 $\Gamma(x, z, u)$ 364
 Q^+ 367
 (A) 372
 (W) 372
 $(\forall)$ 372
 $(\exists)$ 372
 (Cut) 372
rank 373
RPr 375
CFPr 375
RCon 376
CFCon 376
numeral(x) 381